University of Alberta

Residue Molecules: Molecular Representations
and Thermal Reactivity

by

Jeffrey Michael Sheremata

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in
Chemical Engineering

Department of Chemical and Materials Engineering

Edmonton, Alberta
Fall 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-46424-3
Our file Notre référence
ISBN: 978-0-494-46424-3

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The structural chemistry and thermal cracking of an Athabasca bitumen feed were modeled using stochastic methods. The molecular representations were created with a rule based Monte Carlo construction method that represented molecules with a series of aromatic, thioether, and aliphatic groups. After the groups were randomly sampled for a molecule, a connection algorithm linked them together to form intact molecules. Each feed molecule was represented using connection and structural matrices. Large sets containing between 25-300 molecules were constructed and sequential optimization was used to create small subsets of molecules that were consistent with ^{13}C-NMR spectroscopy, ^{1}H-NMR spectroscopy, elemental analysis, vapor pressure osmometry, and simulated distillation data. To accurately represent the analytical data for an Athabasca asphaltene sample, a minimum of five molecules was needed when simulated distillation data was not fit. When simulation distillation data was fit, at least 15 molecules were necessary to create a molecular representation that was consistent with experimental data.

A Monte Carlo approach was used to model the thermal cracking of an Athabasca bitumen feedstock. A molecular representation was created for an Athabasca bitumen feed that contained 17 molecules. Model compound reactivity studies published in the literature were used to determine the probabilities of cracking of various C-C and C-S bonds. These probabilities were used in a continuous reaction algorithm that used matrix transformations to react

feed molecules into product molecules. The reaction simulations were composed of reaction steps. At each reaction step, molecules were stochastically chosen to react and specific bonds were stochastically chosen to crack. Group contribution theory was used to calculate the boiling point of each molecule in the feed and product fractions. The Athabasca bitumen molecular representation was reacted until the simulated liquid product fraction had a composition of <250°C material that was consistent with the experimental data. At this point, most feed molecules had multiple bonds cracked. The composition of 500°C+ liquid product in the thermal cracking simulation was also consistent with the experimental data. The aromaticity, molecular weight, and sulfur content of the cracked liquid product were all consistent with the experimental properties.

Table of Contents

Chapter 1. Introduction	**1**
1.1. Motivation	1
1.2. Background	3
1.3. Research Objectives	4
1.4. References.	5
Chapter 2. Literature Review	**6**
2.1 Introduction	6
2.2 Molecular Representaitons	7
2.2.1. Types of Molecular Representations	7
2.2.2. Quantitative Molecular Representations	8
2.2.3. Quantitative Molecular Representations	8
2.2.3.1. Coal Molecular Representations	9
2.2.3.2. Vector Based Molecular Representations	10
2.2.3.3. Monte Carlo Based Molecular Representations	11
2.2.3.3.1. Neurock's Asphaltene Molecular Representations	11
2.2.3.3.2. Molecular Representations Containing Few Molecules	15
2.3. Molecular Dynamics	18
2.4. Petroleum Origins and Chemistry	19
2.5. Heavy Oil Definitions	21
2.6. Analytical Methods	22
2.6.1 Fractionation	23
2.6.2 Molecular Weight Determination	24
2.6.3. NMR Spectroscopy	25
2.7. Asphaltenes	26
2.7.1. Two Alternative Asphaltene Models: Pericondensed and Archipelago	26
2.7.1.1. The Pericondensed Model	28
2.7.1.1.1. X-ray Diffraction	28
2.7.1.1.2. Fluorescence Depolarization and Mass Spectroscopy	28
2.7.1.1.3. Criticism of Fluorescence Depolarization and Mass Spectroscopy	29
2.7.1.2. The Archipelago Model	31
2.8. Hydrocarbon Pyrolysis and Free Radical Mechanisms	32
2.8.1. Chain Reactions	32
2.8.2. Free Radical Reactions In Petroleum Cracking	35
2.9. Molecular Kinetic Simulation of Thermal Cracking	36
2.9.1. Froment's Molecular Kinetic Thermal Cracking Simulations	37
2.9.2. Dente's Molecular Kinetic Thermal Cracking Simulations	38
2.9.3. Joshi's Molecular Kinetic Thermal Cracking Simulations	38

2.9.4. Monte Carlo Thermal Cracking Simulation 39
2.10. Group Contribution Methods 40
2.10.1. Joback and Reid Group Contribution Method 41
2.10.2. Marrero-Gani Group Contribution Method 42
2.11. Petroleum Model Compounds 43
2.11.1. Model Compound Thermal Cracking Studies 43
2.11.2. Asphaltene Model Compounds 47
2.12. References 48
Chapter 3. Quantitative Molecular Representations of Athabasca Asphaltenes 52
3.1. Background 52
3.2. Objective 52
3.3. Materials and Methods 53
3.3.1. Materials 53
3.3.2. Analytical Data 53
3.3.3. Molecular Generation 59
3.3.4. Optimization 63
3.4. Results 65
3.4.1. Molecular Representation Properties 65
3.4.2.The Effect of Varying the Molecule Population Size 67
3.4.3. The Effect of Varying Molecular Weight 68
3.5. Discussion 72
3.6. Conclusions 76
3.7. References 77
Chapter 4. Quantitative Molecular Representations of Bitumen Residue Fractions 78
4.1. Introduction 78
4.1.1 The Relationship Between Structure and Function 78
4.1.2. Solubility Parameter Modelling of Bitumens 81
4.2. Materials and Methods 82
4.2.1. Residue Molecular Representations 82
4.2.2. Objective Function Formulation 83
4.3. Results and Discussion 85
4.3.1. Molecular Representation Optimizations 87
4.3.2. Molecular Representation Performance 88
4.3.3. Large Scale Optimization Results 92
4.3.4. Example Molecules 93
4.3.5. A Consistent Basis 100
4.3.6. Aromatic Ring Attachments 100
4.3.7. Aromatic Attachments 104
4.4. Conclusions 108
4.5. References 109
Chapter 5. Thermal Cracking Simulations of Bitumen Molecular Representations 110
5.1. Introduction 110
5.2. Data 110

5.3. Thermal Cracking Modeling 117
5.4. Molecular Representations 122
5.4.1. Construction Method 122
5.4.2. Objective Function Formulation 123
5.5. Molecular Representation Results 124
5.6. Thermal Cracking Modeling 141
5.6.1. Model Validation 141
5.6.2. The First Thermal Cracking Simulation Method 141
5.6.3. The Second Thermal Cracking Simulation Method 143
5.7. Discussion 155
5.8. References 161
Chapter 6. Discussion and Recommendations 162
Chapter 7. Conclusions 167
Appendix A. Model Compound Thermal Reaction Studies 169
A.1. Branched Model Compounds 169
A.2.Sulphur Containing Model Compounds 170
A.3. Phenyl Model Compounds 171
A.4. Pyrene Model Compounds 174
A.5. Naphthenic Model Compounds 177
A.6. References 179
Appendix B. The First Molecular Representation Creation Algorithm 180
Appendix C. Molecule Storage and Visualization 187
C.1. Molecule Storage 187
C.2. Molecule Visualization 191
Appendix D. The Modeling of Thermal Cracking Reactions 196
D.1. Incorporation of Model Compound Data 196
D.1.2. Relative Cracking Propensities 197
D.2. Thermal Cracking Modeling 200
D.2.1. The REACT Matrix 200
D.2.2. Thermal Cracking Simulations 202
D.3. The Thermal Cracking Algorithm 205
D.5. References 209
Appendix E. The Second Molecule Construction Algorithm 210

List of Tables

Table 2-1. Attributes used by Neurock *et al.* to create molecular representations. 12
Table 2-2. UNITAR definitions of heavy oils and bitumens. 21
Table 2-3. Properties of light and heavy crude oils. 21
Table 2-4. Bond Dissociation Energies. 35
Table 2-5. Molar yields of products from tridecylcyclohexane thermolysis at 400 ℃ 46
Table 3-1. NMR spectroscopy chemical shifts of proton spectral regions. 55
Table 3-2. NMR spectroscopy chemical shifts of carbon spectral regions. 55
Table 3-3. Definitions of NMR spectroscopy determined chemical groups. 57
Table 3-4. Experimental properties and their respective errors of the Athabasca residue sample. 58
Table 3-5. Comparison between experimental and calculated properties of the asphaltene molecular representation. 66
Table 3-6. Calculated elemental compositions of the six molecules in the asphaltene representation. 67
Table 3-7. Calculated properties of asphaltene representations of different target molecular weights. 71
Table 4-1. Properties of fractionated Athabasca vacuum residue. 82
Table 4-2. Optimized properties of the molecular representations. 85
Table 4-3. Detailed composition of the ten Athabasca residue molecular representations. 86
Table 4-4. Calculated properties of the molecular representations of Fractions 1-3. 89
Table 4-5. Calculated properties of the molecular representations of Fractions 4-6. 90
Table 4-6. Calculated properties of the molecular representations of Fractions 7-9. 91
Table 4-7. Calculated properties of the Fraction 10 molecular representation. 92
Table 5-1. Distillation breakdown, elemental analysis, and group distribution of the Athabasca bitumen feed. 112
Table 5-2. Distillation breakdown, elemental analysis, and group distribution of the liquid Athabasca bitumen thermal cracking products. 113
Table 5-3. Experimental versus calculated properties of the kerosene molecular representation. 127
Table 5-4. Experimental versus calculated properties of the gas oil molecular representation. 128
Table 5-5. Experimental versus calculated properties of the

residue molecular representation. 129
Table 5-6. Experimental versus calculated properties of the *Distributio* feed molecular representation. 130
Table 5-7. Experimental versus calculated properties of the *Summa* feed molecular representation. 131
Table 5-8. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca kerosene naphtha. molecular representation. 132
Table 5-9. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca bitumen gas oil molecular representation. 133
Table 5-10. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca bitumen residue molecular representation. 134
Table 5-11. Elemental compositions, molecular weights, aromaticities, and boiling points of the seventeen molecules in the Athabasca bitumen *Summa* feed representation. 134
Table 5-12. Results from the thermal cracking simulation that used the *Distributio* feed molecular representation. 142
Table 5-13. Gas production, molecular weight, liquid fraction composition, elemental analysis, and carbon type distribution of the descritized *Distributio* molecular representation and the simulated thermal cracking product. 147
Table 5-14. Gas production, molecular weight, liquid fraction composition, elemental analysis, and carbon type distribution of the descritized *Summa* molecular representation and its simulated cracked product. 148
Table 5-15. Molecular weight, sulfur content, aromaticity, and naphthenic values, of both the *Summa* and *Distributio* molecular representation thermal cracking simulations, when the simulated liquid contents of the 500℃+ material ma tched the experimental liquid product data. 149
Table 5-16. Mole balance of a *Summa* molecular representation thermal cracking simulation. 149
Table A-1. Thermal cracking of n-didodecyl sulfide. 171
Table A-2. Molar yields of products from phenyldodecane thermal cracking at 400 ℃. 172
Table A-3. Product selectivities for the thermal cracking of 1,20-di(1-pyrenyl)eicosane at 427 ℃. 175
Table A-4. Molar yields of products from tridecylcyclohexane thermal cracking at 400 ℃. 178
Table A-5. Molar yields of products from 2-Ethyltetralin thermal cracking at 400 ℃. 179
Table E1. Maximum number of attachments allowed to connect to the aromatic groups. 211

List of Figures

Figure 2-1. Representative molecules from a supercritical gas extract of coal at 350°C. 9
Figure 2-2. Representative coal structures created by the CAMSC algorithm. 10
Figure 2-3. A petroleum molecule represented using a single structural vector. 11
Figure 2-4. One of 10,000 molecules in an asphaltene Molecular representation created by a Monte Carlo based algorithm. 12
Figure 2-5. An asphaltene molecule created by a stochastic algorithm. 14
Figure 2-6. Two representative molecules from a Maya Monte Carlo representation proposed by Campbell and Klein. 16
Figure 2-7. Molecules from the Khorasheh *et al.* gas oil representation. 17
Figure 2-8. Pericondensed aspahltene representation of Yen *et al.* 27
Figure 2-9. A general archipelago heavy oil structure. 27
Figure 2-10. Asphaltene molecule proposed by Groenzin and Mullins 29
Figure 2-11. Free reaction chain mechanism. 33
Figure 2-12. Free reaction chain mechanisms involved in hydrocarbon pyrolysis. 33
Figure 2-13. The structures of pristane (a.), phytane (b.), and squalane (c.). 45
Figure 2-14. The structures of didodecyl sulfide. 45
Figure 2-15. The structure of phenyldodecane. 45
Figure 2-16. The structure of 1,20-di(1-pyrenyl)eicosane. 45
Figure 2-17. The Structures of: a. Tridecylcyclohexane. 46
Figure 2-18. The structures of: a. 4,4'-bis-(2-pyren-1-yl-ethyl)-[2,2']bipyridinyl (PBP); b. Hexabenzocoronene. 48
Figure 3-1. Various aliphatic and aromatic carbons present in the asphaltene molecular representation. 56
Figure 3-2. Flowchart for construction of a molecular representation using the Monte Carlo construction method. 60
Figure 3-3. The aromatic groups present in the asphaltene representation. 61
Figure 3-4. $C_{283}H_{337}N_3O_4S_9$. 68
Figure 3-5. $C_{230}H_{302}N_4O_2S_{10}$.Wt.: 3476 g/mol. 69
Figure 3-6. $C_{318}H_{395}N_6O_6S_8V$ Mol. Wt.: 4705 g/mol. 70
Figure 3-7. Sequential optimizations of populations of 25, 50, 100, and 150 asphaltene molecules. 71
Figure 4-1. Sequential optimization for the creation of a molecular representation of Fraction 10. 87
Figure 4-2. Large scale sequential optimization of Fraction 8. 93
Figure 4-3. Solubility parameters of the ten residue fractions. 95
Figure 4-4. Schematic of two molecules from Fraction 1. 96
Figure 4-5. Schematic of a molecule from Fraction 4. 97

Figure 4-6. Schematic of a molecule from Fraction 7. 98
Figure 4-7. Schematic of an asphaltene-like molecule from Fraction 10. 99
Figure 4-8. Average number of aromatic rings/1000 C. 101
Figure 4-9. Average number of aromatic rings/aromatic cluster. 102
Figure 4-10. Average distance between aromatic clusters. 105
Figure 4-11. Average length of aromatic terminal attachments. 106
Figure 4-12. Average number of aromatic terminal attachments. 107
Figure 5-1. Simulated distillation curve for the entire Athabasca bitumen feed and liquid product. 115
Figure 5-2. Simulated distillation curve for the Athabasca bitumen naphtha, gas oil and residue fractions. 115
Figure 5-3. Group assignments obtained from the modified NMR spectroscopic heavy oil analysis method. 116
Figure 5-4. Additional group assignments obtained from the modified NMR spectroscopic heavy oil analysis method. 116
Figure 5-5. The structures of pristane (a.), phytane (b.), and squalane (c.). 117
Figure 5-6. The structures of didodecyl sulfide 117
Figure 5-7. The structure of phenyldodecane. 118
Figure 5-8. The structure of 1,20-di(1-pyrenyl)eicosane. 118
Figure 5-9. The Structures of: a. Tridecylcyclohexane. 118
Figure 5-10. Relative cracking propensities of various C-C and C-S bonds in a variety of model compounds. 119
Figure 5-11. Overview of the thermal cracking algorithm. 120
Figure 5-12. Transformations and reactions modeled in the thermal cracking simulations. 121
Figure 5-13. Molecular representation construction algorithm flowchart. 122
Figure 5-14. A plot of the residue molecular representation objective function as a function of number of molecules included in the molecular representation. 126
Figure 5-15. Boiling point distribution of the Athabasca bitumen naphtha molecular representation plotted against experimental simulated distillation data. 135
Figure 5-16. Boiling point distribution of the Athabasca bitumen gas oil molecular representation plotted against experimental simulated distillation data. 135
Figure 5-17. Boiling point distribution of the Athabasca bitumen residue molecular representation plotted against experimental simulated distillation data. 136
Figure 5-18. Boiling point distribution of the *Distributio* Athabasca bitumen feed molecular representation plotted against experimental simulated distillation data of the entire feed. 136
Figure 5-19. Boiling point distribution of the *Summa* Athabasca bitumen feed molecular representation plotted against experimental simulated distillation data of the entire feed. 137

Figure 5-20. Three molecules from the Athabasca bitumen naphtha molecular representation. 137
Figure 5-21. Three molecules from the Athabasca bitumen gas oil molecular representation. 138
Figure 5-22. Athabasca bitumen residue molecular representation molecule (a). 138
Figure 5-23. Athabasca bitumen residue molecular representation molecule (b). 139
Figure 5-24. Athabasca bitumen residue molecular representation molecule (c). 139
Figure 5-25. Athabasca bitumen residue molecular representation molecule (d). 140
Figure 5-26. Conversion of the 500℃ + material du ring a thermal cracking simulation that used the *Summa* molecular representation. 150
Figure 5-27. Gas production and liquid fraction composition observed during a thermal cracking simulation that used the *Summa* molecular representation. 150
Figure 5-28. Changes in the liquid fraction mass content of 500℃ material during a thermal cracking simulation s that used the *Summa* and *Distributio* molecular representations. 151
Figure 5-29. Changes in the liquid fraction average molecular weight during visbreaking simulations that used the *Summa* and *Distributio* molecular representations. 151
Figure 5-30. Changes in the liquid fraction aromatic and naphthenic contents during thermal cracking simulations that used the *Summa* and *Distributio* molecular representations. 152
Figure 5-31. Changes in the liquid fraction sulfur composition during thermal cracking simulations that used the *Summa* and *Distributio* molecular representations. 152
Figure 5-32. Boiling point distribution of liquid fraction molecules of a simulated *Summa* cracked product plotted versus the experimental SIMDIS data. 153
Figure 5-33. Cumulative distribution of molecules in the liquid product fraction of the *Summa* molecular representation during a thermal cracking simulation.characterized by the number of bonds cracked/feed molecule (in different reaction steps). 154
Figure A-1. The structures of pristane (a.), phytane (b.), and squalane (c.). 169
Figure A-2. The structures of didodecyl sulfide (a.) and 1-dodecanethiol (b.). 170
Figure A-3. The structure of phenyldodecane. 171
Figure A-4. The structures of: a. pentadecylbenzene and b. 2-n-pentadecylpyridine. 173
Figure A-5. The structure of 1,20-di(1-pyrenyl)eicosane. 174
Figure A-6. The structure of 1-dodecylpyrene. 176
Figure A-7. The structures of: a. 2-(3-phenylpropyl)

naphthalene. b. 1,3-bis(1-pyrene) propane. 176
Figure A-8. The Structures of: a. Tridecylcyclohexane and
b. 2-Ethyltetralin. 177
Figure C-1. The matrix representation of a small hydrocarbon molecule. 188
Figure C-2. A more advanced matrix representation of a
small hydrocarbon molecule. 190
Figure C-3. A MOL file for a simple hydrocarbon molecule 193
Figure D-1. The structure of a. phenynonane b. phenyldodecane. 197
Figure D-2. Relative cracking propensities of various C-C and
C-S bonds in a variety of model compounds. 199
Figure D-3. The REACT matrix. 201
Figure D-4. Transformations and reactions modeled in the
thermal cracking simulations. 203
Figure D-5. Overview of the thermal cracking algorithm. 204

Nomenclature

Stochastic	Weighted random sampling
PDF	Probability distribution function

NMR Spectroscopy	Nuclear Magnetic Resonance Spectroscopy
Aromaticity	Percent of total carbon that is aromatic
Q_1	Alkyl-substituted aromatic quaternary carbon
Q_2	Bridgehead aromatic quaternary carbon
C_1	Aromatic CH beside a Q_2 carbon
C_2	All aromatic CH that are not a C_1 carbon
α-CH_3.	Terminal CH3 attached either to an aromatic, naphthenic ring, or CH branch
β-CH_3	Terminal CH3 on an aromatic side chain with two carbons
γ-CH_3	Terminal CH3 on a long chain paraffin chain

δ	Solubility parameter
MW	Molecular weight
H/C	Hydrogen to carbon ratio
IBP	Initial boiling point

Summa	Molecular representation based on data for an entire feed
Distributio	Molecular representation based on individual representations for various fractions

RCP	Relative cracking propensity
ACP	Absolute cracking propensity
REACT	Thermal cracking matrix
J	Structural matrix
CON	Connection matrix
MOL	A computer file that stores molecules

Chapter 1. Introduction

1.1.. Motivation

In Science and Engineering, molecular chemistry is frequently studied. After the structural chemistry of a molecule is known, it often becomes possible to optimize and enhance the properties and performance of environments or systems that contain the molecule of interest. For example, molecular biologists study the structure and function of deoxyribonucleic acids (DNAs) and ribonucleic acids (RNAs). When one has knowledge of the structure of a macromolecule, one can infer its function. James Watson and Francis Crick were able to deduce how DNA was transcribed into RNA and translated into proteins by solving the structure of DNA.[1] Thus, by knowing the structure of DNA, the function of DNA was deduced. Furthermore, the structural work of Watson and Crick paved the way for the genetic engineering of DNA. Once the structure of DNA was known, it became feasible to modify the structure and function of DNA. Thus, when the structure of a molecule is known, it becomes possible to engineer or enhance a molecular system to achieve an optimal design of the overall system.

There are several reasons why it is necessary to study the structure of heavy oils. By improving our knowledge of the structure of heavy oil, current recovery, extraction, and refining technologies will be improved and new technologies will be developed. In science, some problems are best solved at the micro level while others are best solved at the macro scale. However, it is common to obtain insights about an overall system by studying individual organization at the micro level. For example, in economics a microeconomist might study individual buying and selling markets. However, the economist might be capable of predicting certain economic trends for an entire country based upon the study of certain microeconomic level markets. A fluid mechanics scientist might study fluid flow at the micro level and gain insights about how the fluid would flow in a larger system. The structural chemistry of individual

molecules ultimately determines and influences the physical properties of petroleum at the bulk scale. Thus, information at the structural (micro) level will give us insights into how petroleum behaves at the bulk (macro) level. Also, new structural information will allow engineers to solve a variety of problems associated with petroleum and petroleum processing that can only be solved or calculated at the micro scale. Detailed structural information allows engineers to make representations of petroleum at the molecular level. When engineers have information at the molecular level, they can estimate physical properties and calculate reaction rates for both individual molecules and bulk petroleum mixtures. This will enable engineers to use the structural information to optimize current processes and create new ones. Thus, structural chemistry will have benefits to engineers at both the micro and macro levels of engineering.

Chemical structure ultimately determines how molecules react. In the case of petroleum molecules, the chemical structure of a petroleum molecule will determine how it will break apart when it is thermally processed. With knowledge of petroleum structure, engineers will more accurately be able to predict how petroleum molecules break apart when they are processed. They will be more capable of constructing models that will predict the properties and products of a petroleum processing operation using a given feed. Also, given a feed, engineers will be more capable of choosing processes and developing new technologies to process it when the structural chemistry is known.

Computer simulation has a very important role in the engineering of heavy oil. Reservoir engineers utilize simulators to predict how much heavy oil can be recovered from a deposit and the rate of recovery. Reservoir engineers calculate the profitability associated with different heavy oil deposits and the relative risks. Heavy oil reservoir simulators use models of the heavy oils they are simulating. Insights into the structural chemistry of heavy oil will improve the reservoir simulation models that reservoir engineers use. This will allow them to more accurately model fluid flow in heavy oil reservoirs. As a result, reservoir

engineers will generate more precise economic forecasts and more accurate risk assessments. Likewise, process and design engineers use design programs to design upgraders and refineries, and to optimize the operating conditions. As our knowledge of the chemical structure of heavy oil increases, our knowledge of what happens during the refining and upgrading of heavy oil improves as well. Thus, process engineers will be able to design better upgraders and refineries as knowledge of heavy oil structure increases. Design software also requires heavy oil models. With an improved understanding of heavy oil structure, design software will be more effective an optimizing process conditions and sizing equipment. Thus, improved knowledge of heavy oil structure will allow process engineers to both create better refining and upgrading processes, and be more capable of choosing operating and design criteria that will optimize process economics.

1.2. Background

The chemical complexity and diversity of petroleum asphaltenes and residues make analyzing and modeling their structural chemistry very challenging. A major complication in the study of heavy oil structural chemistry is that multiple experimental techniques must be used to generate information about its molecular nature. Thus, mathematics, statistics, and simulation are required to create heavy oil structural models that are consistent with analytical data. The technique that accomplishes this is quantitative molecular representation.

Quantitative molecular representations integrate information from various analytical techniques to yield an overall picture of a heavy oil fraction at the molecular level. Quantitative molecular representations have been created using Monte Carlo methods based on data from ^{1}H-NMR spectroscopy, elemental analysis, and VPO for various asphaltenes and residues.[2,3] The number of

molecules contained in published representations ranges from 12 to 10,000 molecules. Quantitative molecular representations can also be used to generate highly versatile kinetic and reactivity models and to calculate properties such as density, boiling point and solubility.[1,4,5]

1.3. Research Objectives

The overall goal of this research was to simulate the thermal cracking of heavy oil using quantitative molecular representations. Thus, the two major goals of this project were:

1. Create quantitative molecular representations of heavy oil fractions using a minimum number of molecules. The objectives of this stage were:
    a. Develop a computationally efficient matrix framework that efficiently represents and stores heavy oil molecules.
    b. Create a stochastic algorithm that builds heavy oil molecules consistent with analytical data.
    c. Create a visualization algorithm to view the molecules in the molecular representations.
    d. Develop mathematical optimization routines were that create small populations of molecules consistent with analytical data.

2. Create an algorithm that simulates the thermal cracking of heavy oil molecular representations.
    a. Develop a theoretically reasonable approach to stochastically model the thermal cracking of bitumens using literature data from thermal cracking model compound studies.
    b. Identify model compound reactivity studies in the literature that are applicable to the thermal cracking of heavy oil. Incorporate the data from these studies into the thermal cracking model.

c. Develop and create the matrix, stochastic, numerical, and mathematical techniques necessary to simulate the chemical reactions that occur during the thermal processing of heavy oil.
d. Assure that the algorithm does a reasonable job of simulating the thermal cracking of bitumen molecular representations.

1.5. References

1. Watson, J.D.; Crick, F.H.C. *Nature*. **1953**, *171*, 737-738.
2. Neurock, M.; Nigam, A.; Trauth, D.; Klein, M.T. *Chem. Eng. Sci.* **1994**, *49*, 4153-4177.
3. Campbell, D M.; Klein, M.T. *Applied Cat. A: General* **1997**, *160*, 41-54.
4. Joshi, P. Klein, M.T. *Rev. Process. Chem. and Eng.* **1998**, *1*, 111-140.
5. Rogel, E. *Energy Fuels*. **1997**, *11*, 920-925.

Chapter 2. Literature Review

2.1. Introduction

The goal of this research was to create stochastic thermal cracking models of bitumen thermal cracking using molecular representations. The first part of this literature review focuses on elements that are important to creating molecular representations. Molecular representations have a fundamental role in this thesis. A detailed history of coal and petroleum molecular representations will first be presented. Petroleum is created from biological source material. The relevance of source material chemistry to molecular representations will then briefly be discussed. Molecular representations require many types of experimental data. The analytical techniques necessary to generate data to create molecular representations will also be discussed. Currently, there are two different views of asphaltene molecular structure: pericondensed and archipelago. It is important to choose a structural framework when molecular representations are created. Thus, both pericondensed and archipelago representations of asphaltenes will be reviewed.

The second part of this literature will focus on literature that was important to the development of the thermal cracking algorithm. Free radical chain reactions are responsible for the cracking that occurs during bitumen upgrading. Free radical reactions will briefly be reviewed. Model compounds have been used by petroleum scientists to gain insight into how specific functional groups react during thermal processing. Information from model compound reactivity studies was used in the thermal cracking algorithm used in this thesis, so these studies will be discussed. Previous thermal processing simulations will also be summarized in the literature review. Group contribution theory is also discussed in the literature review. Group contribution theory allows physical property calculation to be calculated from molecular representations. In thermal cracking

simulations, it is necessary to calculate various physical properties of simulated feeds and reactants such as boiling points.

2.2. Molecular Representations

2.2.1. Types of Molecular Representations

The most general definition of petroleum molecular representations is "illustrations of petroleum molecules." There are two types of petroleum molecular representations - schematic and quantitative. Both types of molecular representations are interpretations of analytical data. Schematic representations are manually sketched and are usually single molecules used to illustrate trends in the data. The schematic (or average molecule) approach of interpreting petroleum analytical data began in the early 1960s.[1,2] This approach uses correlations that rely on ^{1}H-NMR spectroscopy data to calculate average properties, including aromaticity and degree of substitution.[1,2]

Quantitative representations are constructed by computers and use mathematical methodologies to create series of molecules consistent with the analytical data on which they are based. Most quantitative molecular representations involve Monte Carlo simulation. A major constraint is that molecules must be statistically and mathematically consistent with the experimental data that they are based on. Molecular representations must be robust and they must be capable of incorporating new experimental insights into their framework.

2.2.2. Schematic Molecular Representations

The first published (the "Yen" representation) asphaltene representation appeared in 1972 (Figure 1-4).[3] The "Yen" representation is significant because it was the first published asphaltene representation and it was the first heavy oil representation to be based on analytical data. The representation was based on ^{1}H-NMR spectroscopy, X-ray diffraction, and ring compactness data for a Langunillas oil. The aromatic, naphthenic, methylene, and methyl proton compositions of the structure were very close to the experimentally determined values. This representation strongly influenced the way that many asphaltene scientists and engineers perceive asphaltene molecules. More than 30 years after it was published, the influences of this representation are still observed in asphaltene literature. This representation was schematic in the sense that it was drawn to represent the overall trends displayed in the data.

2. 2.3. Quantitative Molecular Representations

Schematic molecular representations do not accurately reflect the analytical data that they are based on. However, by definition quantitative representations are consistent with the analytical data from which they were created. Mathematical, graphical and statistical techniques are thus necessary to transform the data generated by analytical techniques into the molecular form. The next few subsections discuss the evolution of quantitative molecular representations.

2.2.3.1. Coal Molecular Representations

The historical precursors to petroleum molecular representations are the coal molecular representations which were the first type of representations to be created using a mathematical methodology. Bartle *et al.* analyzed a coal supercritical gas extract using a variety of experimental techniques including ^{1}H-NMR spectroscopy, ^{13}C-NMR spectroscopy, and infrared mass spectrometry.[4] Model compounds (Figure 2-1) were proposed that had properties similar to the analytical data. A multidimensional graphical approach was used to compare the analytical coal data to the model compound data. The model compounds that best satisfied the analytical data were selected. A single molecule was able to adequately represent a high molecular weight coal fraction. Conversely, series of molecules were necessary to represent low molecular weight coal fractions.

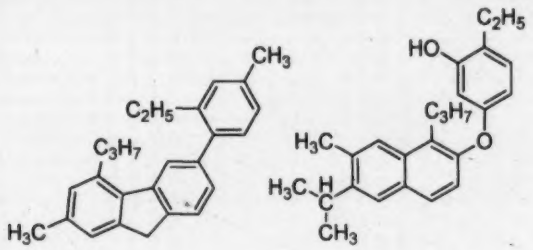


Figure 2-1. Representative molecules from a supercritical-gas extract of coal at 350°C.[4]

The first molecular representations created entirely by a mathematical framework were those created by Oja *et al.*[5] They used an algorithm called CAMSC (computer assisted molecular structure construction) to create coal molecules that satisfy NMR spectroscopy, elemental analysis, and molecular weight data. This algorithm made molecules by creating different permutations of various aromatic, naphthenic, and aliphatic structural groups. The algorithm

was used to create molecules both for a purified compound from a coal extract and for a coal distillate (Figure 2-2).

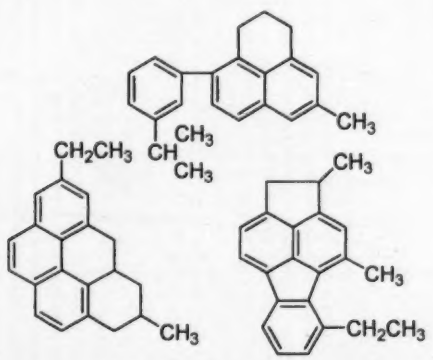
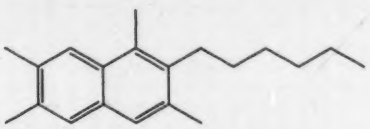


Figure 2-2. Representative coal structures created by the CAMSC algorithm.[5]

2.2.3.2. Vector Based Molecular Representations

Jaffe *et al.* developed an efficient vector framework that created molecular representations for both residues and smaller petroleum fractions.[6,7] This approach is termed "structure oriented lumping" and it represents petroleum molecules as composed of various groups such as: benzene, multi-ring aromatics, naphthenics, alky chains, etc. For simple molecules, a single vector documents all of the chemical groups that compose the molecule (Figure 2-3). The vector does not document the exact chemical arrangements within a molecule. Larger residue molecules are composed of multiple vectors corresponding to different cores. The number of carbons that connect each core is documented. A connection table is also required to document the interconnections of all of the various cores. The advantages of using vectors include: they are very computationally efficient, they are easy to manipulate, and they are straightforward to analyze. This vector approach also simplifies the modeling of the reactions associated with petroleum processing. Mathematical

operations are associated with various reactions such as: aromatic saturation, aromatic ring opening, paraffin isomerisation, and cracking. Rules determine both what groups and cores can participate in specific reactions and the products and vectors that result from a reaction.



A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	IH	AA	NS	RS	AN	NN	RN	NO	RO	KO
1	1	0	0	0	0	0	0	0	10	0	4	0	0	0	0	0	0	0	0	0	0

Figure 2-3. A petroleum molecule represented using a single structural vector.[6] A6 represents a six carbon aromatic ring. A4 represents a four carbon aromatic ring implement that is attached to A6. R represents the total number of alkyl carbons associated with the aromatic or aliphatic compounds.

2.2.3.3. Monte Carlo Based Molecular Representations

2.2.3.3.1. Neurock's Asphaltene Molecular Representations

Neurock *et al.* used a Monte Carlo methodology to construct an asphaltene molecular representation based on data from ^{1}H-NMR spectroscopy, elemental analysis, and VPO (Figure 2-4).[8] This was the first published study of a quantitative heavy oil molecular representation. The structural parameters that were both calculated from the analytical data and used to construct the molecular representation included: the number of unit aromatic sheets, number of aromatic rings, arrangement of the aromatic cores, number of naphthenic rings, the degree of substitution, and the length of sidechains. Normal probability distributions were fitted to each of these groups and randomly sampled. Each molecule was represented as a unit sheet of aromatic and naphthenic rings, with both aliphatic sidechains and thioether attachments. The asphaltene fraction

was represented by 10,000 molecules. The asphaltene representation was used in a thermolysis reaction simulation.

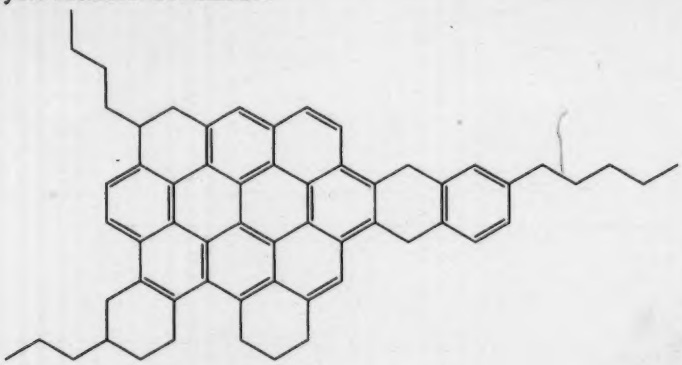


Figure 2-4. One of 10,000 molecules in an asphaltene molecular representation created by a Monte Carlo based algorithm.[8]

Neurock *et al.* created molecular representations for both an offshore California asphaltene and a Kern River heavy oil[9]. In this study, petroleum molecules were described as being a collection of molecular attributes. The attributes used by Neurock are presented in Table 2-1.

Table 2-1. Attributes used by Neurock *et al.* to create molecular representations.[9]

Aromatic	Number of unit sheets/molecule
	Number of aromatic rings/unit sheet
	Configuration of aromatic core
	Degree of substitution off aromatic carbon
Naphthenic	Number of naphthenic rings/unit sheet
	Degree of substitution off naphthenic carbon
Paraffinic	Length of aliphatic side chain attached to aromatics
	Length of each sidechain attached to naphthenics
	Carbon number

Each aromatic and asphaltene molecule in this study was represented as a collection of unit sheets of aromatics and naphthenic rings. The unit sheets were connected to each other by aliphatic bridges. The unit sheets also contained aliphatic side chains. Paraffins were described by carbon number. Naphthenics were described by the number of naphthenic rings, percent of carbons substituted, and substituted carbon chain length. Random sampling from Gaussian and gamma functions (probability distribution functions or PDFs) generated the number of each attribute for each molecule. The specific arrangement of carbons around each aromatic unit sheet was also randomly determined. A detailed thermodynamic algorithm was used to calculate boiling point and solubility distribution of the molecular representations.

For the offshore California asphaltene (Figure 2-5), excellent agreement was obtained between the predicted properties of the molecular representation and the experimental data. The average predicted molecular weight, 2,770 g/mol (normal PDF), 2,726 (gamma PDF) deviated slightly from the experimentally determined value (2,683 g/mol). The calculated simulated distillation curve for the 10,000 Kern River molecules was in very good agreement with the experimental simulated distillation curve - the fit between the two had a standard deviation of 20.8 °C.

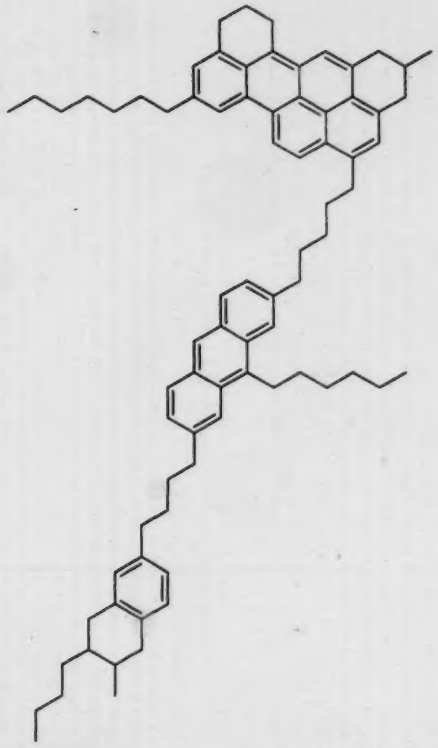


Figure 2-5. An asphaltene molecule created by a stochastic algorithm.[9]

2.2.3.3.2. Molecular Representations Containing Few Molecules

Campbell and Klein developed a Monte Carlo quadrature method that represented petroleum fractions using a limited number of molecules.[10] Unlike other methods that use thousands of molecules to represent a fraction, the method published by Campbell and Klein represented petroleum mixtures using 10-100 molecules. This technique is an extension of the Neurock *et al.* method: that represented molecules as a series of chemical groups randomly sampled from PDFs. In this study, paraffin molecules were classified by their lengths; naphthenics - both aromatic and aliphatic - were classified by the number of rings they contain, the number of sidechains they have, and the length of their

sidechains. Aromatics were classified by the number of sidechains they have, and the length of their sidechains, and asphaltenes were classified by the number of rings in their unit sheets. Each molecule was defined as a collection of each of these attributes. A sum of squares objective function compared the experimentally determined versus predicted properties of the molecular representation for molecular weight, C/H ratio, aliphatic proton content, aromatic proton content, SARA distribution, and boiling distribution. Each attribute was assigned a gamma PDF. The value of the objective function was the sum of the squared weighted differences of all properties calculated from all of the molecules in a molecular representation. Global simulated annealing was used to adjust the α and β parameters of each PDF, so that the objective function was minimized. Global simulated annealing yields molecular representations that contain thousands of molecules.

Quadrature sampling was used to produce molecular representations that contain small numbers of molecules. Each PDF was divided into two, three, or four sections. Each section of each PDF was randomly sampled to create a small population of 30-40 molecules. This technique assured that the entire distribution was sampled. Numerical optimization was then used to change the mole fraction of each molecule to minimize the objective function for the petroleum fraction being modeled. Molecular representations were created for Arabian light (molecular weight of 842 g/mol), a Mayan residue (molecular weight of 944 g/mol), and naphtha (molecular weight of 101 g/mol) petroleum fractions.

For the Arabian light and Mayan residue (Figure 2-6), the predicted boiling point distributions, the H/C ratios, the sulphur compositions, and the aromatic proton compositions all closely matched the experimental data. The naphtha molecular representation contained approximately 20 molecules. About half the initial molecules were assigned "0" mole fractions. For the naphtha fraction, Klein compared the molecules that his algorithm predicted to those that have been experimentally identified in naphtha. Most of the molecules this Monte

Carlo algorithm generated have been experimentally identified in naphtha. However, the predicted paraffin distribution in the molecular representation deviated significantly relative to the experimentally determined distribution. The simulated distillation curve of the naphtha molecular representation closely matched the experimental boiling point curve.

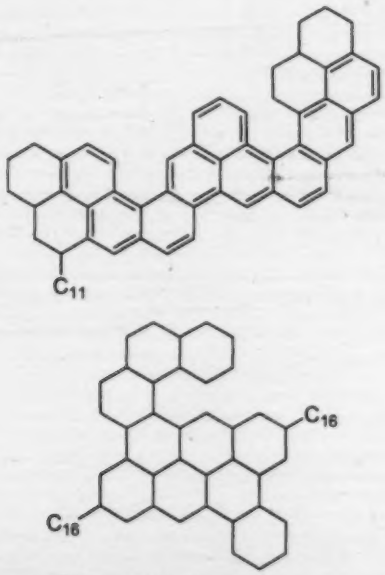


Figure 2-6. Two representative molecules from a Maya Monte Carlo representation proposed by Campbell and Klein.[10]

Khorasheh *et al.* developed a versatile approach to represent high molecular weight petroleum molecules.[11] A matrix framework was used that allowed nonlinear molecules to be constructed. The groups that compose the architectural framework of a nonlinear molecule are arranged in a nonlinear – branched or irregular – path. Most molecular representations contain all of the groups that compose a molecule in a linear arrangement – all of the groups that compose the architectural framework of the molecule are arranged in a linear path.

An Alberta crude oil with an average molecular weight of 370 g/mol was modeled in this study (Figure 2-7). ^{1}H and ^{13}C-NMR spectroscopy were used to analyze the petroleum sample. 31 chemical groups including aromatic rings, naphthenic rings, aliphatic carbons and thioether bridges were used to construct the molecular representation. Nonlinear optimization was used to calculate the concentration of each group: linear constraints based on ^{1}H-NMR spectrometry data and a quadratic objective function based on ^{13}C-NMR spectrometry data were used.

A molecule was represented using a connectivity matrix, an atom identification vector, and a connection vector. The connectivity matrix documents the connection of each group to each and every other group. The chemical groups were randomly connected together to create molecules. 600 molecules were constructed and the concentrations of the groups in the molecular representations were compared to the group concentrations obtained from the constrained optimization. Averaged over the 600 molecules, a 5.6% error was obtained for each group.

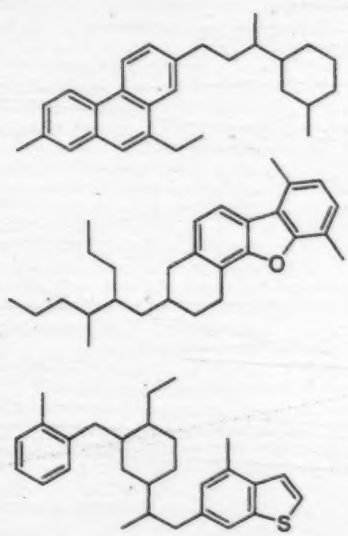


Figure 2-7. Molecules from the Khorasheh *et al.* a gas oil molecular representation.[11]

2.3. Molecular Dynamics

Molecular representations can be used in molecular dynamics simulations. They can be used to study how individual molecules fold or interact internally. Alternatively, they can see how populations of molecules interact both amongst themselves and with other classes of molecules such as solvents.

Modern computational power has fostered the development of molecular dynamics. Molecular dynamics uses energy and force balances on individual atoms in a molecule. Molecular mechanics algorithms gradually adjust the spatial positions of atoms in a molecule to achieve an overall minimal energy, three-dimensional confirmation. The biotechnology and pharmaceutical industries were the first to utilize molecular dynamics. They used this technique to study drug-receptor interactions. Molecular mechanics is now proving to be a useful technique to study the chemical structures of heavy oils, residues, and asphaltenes. The interaction between different molecules can also be studied. Forces and bonds that stabilize molecules and associations between molecules can be identified by molecular mechanics. These forces includes: Π-Π bonding, hydrogen bonding, van der Walls interactions, electrostatic, charge transfer, and induction.[12]

Murgich has used molecular mechanics to study how asphaltenes interact with solvents and other heavy oil molecules. Both pericondensed and archipelago asphaltenes have been studied using molecular dynamics by Murgich.[13] Murgich found that pericondensed asphaltenes associated by stacking interactions between the large aromatic sheets. It was also found that alkyl and cycloalkyl groups attached to the aromatics interfered with the stacking interactions. Conversely, the association of archipelago molecules was more complicated: in addition to aromatic-aromatic associations, bridging and hydrogen bonding results in tangling and bending of the molecules.

The interactions between resins and asphaltenes have also been studied by Murgich.[14,15] Murgich found that asphaltenes have a high degree of selectivity for resins of their own crude. Resins preferentially interact with aromatic rich adsorption sites that are on the outer surface of the multidimensional asphaltene molecules. The corresponding resins of an asphaltene possess a similar three dimensional structure and, therefore, have a very tight fit with the asphaltenes. When, the resins are not structurally similar to the asphaltenes, the interaction energy significantly decreases and binding does not occur as readily. This might explain why resins from one oil are sometimes unable to solubilize asphaltenes from other oils. Likwise, Murgich has shown that Athabasca asphaltenes preferentially interact with their resins than with solvent molecules such as toluene and n-octane.[16]

2.4. Petroleum Origins and Chemistry

The first formal theory regarding the origins of petroleum was postulated by Mendeleyev and Berthelot in the late eighteenth century.[17,18] Mendeleyev and Berthelot proposed that petroleum was derived from inorganic source material and formed deep within the earth. The biological origin of petroleum formation was first proposed by Engler in 1884.[19] In 1934, Triebs found that petroleum both exhibited optical activity and contained porphyrins.[20] Triebs deduced that porphyrins were the transformation products of chlorophyll. Optical activity and chlorophyll are both fingerprints of biological systems and together they provide overwhelming support for the biological origin of petroleum formation.

Petroleum is formed from a wide variety of biological molecules.[20] Hydrocarbons are created from fatty acids when their carboxyl groups are removed. Pigments, fatty acids, sterols, and amino acids are all source

molecules for the components found in petroleum. Many of the chemical groups found in petroleum are biomarkers.[20] Biological and thermal transformations gradually react and transform chemicals from the biological source material – marine phytoplankton – into slightly different molecules. The resulting transformed molecules are termed biomarkers. These molecules retain most of the chemical structure of the original starting material.

Scientists and engineers possess detailed knowledge about the structural biochemistry of the phytoplankton source material that petroleum originates from. The families of reactions involved in transforming the source molecules into petroleum have been identified. These reactions are relatively mild and are predominantly biological reactions that are catalyzed by enzymes: decarboxylations, deaminations, cyclizations, hydrogenations, and isomerizations.[20] The high pressure and temperature associated with petroleum deposits result in light cracking. Thus, on a chemical level there are not many differences between the biological petroleum source material and the final petroleum product. All proposed structures of petroleum and petroleum chemical groups should be consistent with the chemistry of the biological source material.

All petroleum, including heavy oils and bitumens, are composed of three principal classes of hydrocarbon groups: the paraffins (general formula C_nH_{2n+2}), the naphthenics – cycloalkanes - and the aromatics. Paraffins are saturated hydrocarbons that can either be linear or branched. However, parrafins can not contain any ring structures. Naphthenics are saturated hydrocarbon ring systems that contain one or more rings which may have paraffin side chains. Aromatics are hydrocarbons that contain one or more aromatic nuclei. The aromatics might be fused together by biphenyl linkages or connected together by diphenyl methane linkages. Likewise, aromatics can be fused with naphthenic rings and/or paraffin side chains. In addition to the parrafins, aromatics, and naphthenics, petroleum and heavy oils contain organic sulfur groups such as

sulfides and cyclic sulfides. Some aromatics contain heteroatoms such as sulfur, nitrogen, and oxygen.

2.5. Heavy Oil Definitions

Viscosity and density (Table 2-2) are both used to define bitumens and heavy oils.[21] The density of bitumen is approximately 1,000 kg/m^3. The UNITAR definitions of bitumen and heavy oil are found in Table 1-1. Bitumens and heavy oils have greater densities, viscosities, and heteroatom compositions than light crude oils (Table 2-3). On a weight basis, bitumens are composed of at least 50% of residue material with a boiling point greater than 523 °C.

Table 2-2. UNITAR definitions of heavy oils and bitumens.[21]

	Viscosity (mPa•s)	Density (g/cm^3)	API Gravity*
Heavy Oil	10^2-10^5	0.934-1.0	20-10
Bitumen	$>10^5$	>1.00	< 10

*API gravity=141.5/(specific gravity at 15.6°C) – 131.5

Table 2-3. Properties of light and heavy crude oils.[21]

	Light Crude	Cold Lake	Athabasca	Morichal
API Gravity	38	10	9	4.9
Sulphur (wt%)	0.5	4.4	4.9	4.1
Nitrogen (wt%)	0.1	0.4	0.5	0.8
Metals (PPM)	22	220	280	863
Viscosity ($m^2/s*10^6$ at 40 °C)	5	5000	7000	
Vacuum Resid 525 °C+,(Liquid Vol %)	11	52	52	80

Many heavy oil researchers have focused on the asphaltenes which are responsible for many of the problems and complications associated with heavy oil extraction, transportation, and processing. Asphaltenes are defined by solubility: soluble in benzene but insoluble in *n*-pentane. Asphaltenes are aromatic, have the highest molecular weights of all the molecules present in heavy oil, and are characterized by high heteroatom contents[22]. Asphaltenes comprise between 15-20% of Athabasca bitumen and contain sulphur (8-9%), oxygen (2-3%) and nitrogen (1%).[23]

2.6. Analytical Methods

Heavy oil structural chemistry research has been at the forefront of technological innovation since the 1960s. As a result of the tremendous complexity associated with heavy oil structural chemistry, researchers have had to rely on advanced technology to gain insights into the structural chemistry of heavy oils. Ultimately, the structural determination of heavy oil is an unconstrained problem that has a high degree of ambiguity. The complex nature of heavy oil chemistry makes it necessary to use technologically advanced analytical methods to gain insights. However, the complex nature of the analytical methods makes it possible to make unreasonable conclusions about heavy oil structural chemistry.

Heavy oil researchers have a history of making unreasonable conclusions based upon data obtained from technologically advanced but unproven techniques. New insights into heavy oil structural chemistry often occur with the development of new technologies. However, researchers and scientists must utilize proper experimental controls to prevent unreasonable conclusions from being made. Likewise, it is critically important to understand the scientific fundamentals associated with any new technology. When the fundamentals are not understood or appreciated, poor and misleading conclusions occur. Even when heavy oil scientists and researchers have innovative and new technology

available to them, they still must utilize logic, common sense, and the scientific method. There will always be experimental results that contradict other preexisting experimental results. However, there are times when the contradictions represented by certain conclusions are so illogical, that they should never be proposed. An example of this is when a conclusion contradicts basic scientific fundamentals. An extreme example of this is a conclusion that contradicts conservation of energy or mass laws.

Heavy oils and bitumens are very complicated mixtures that contain hundreds of thousands of unique molecules. The diverse elemental compositions of heavy oil molecules, combined with the chemical complexities, and the enormous number of unique molecules found in a heavy oil fraction, renders heavy oil a challenging mixture to experimentally characterize. As a result of the chemical complexity, many different techniques must be used to study heavy oil chemical structure.

2.6.1 Fractionation

Heavy oils are usually separated into fractions that are easier to analyze and characterize. Distillation can fractionate a heavy oil sample into multiple boiling point cuts. However, conventional distillation cannot resolve cuts with boiling points above 524 °C. Alternatively, adsorption and solubility can be used to separate a heavy oil fraction. SARA first utilizes solvent precipitation to precipitate out the asphaltenes of a heavy oil sample using n-heptane. Adsorption chromatography is then used to separate the remaining material into resin, saturate, and aromatic fractions.

Supercritical fractionation and extraction can be used to prepare narrow cuts from a residue or heavy oil fraction.[24] In supercritical fractionation, heavy oil is extracted with a solvent at a temperature and pressure above its thermodynamic critical point. The pressure in a supercritical fractionation unit is

gradually increased to yield different fractions. This type of fractionation method is ideal for separating the material in the + 524 °C fraction that is obtained from conventional distillation. Chung *et al.* were able to separate a residue derived from an Athabasca bitumen into ten different fractions using supercritical extraction with n-pentane as the solvent.[24] In the first nine fractions, the molecular weight gradually changed from 500 to 1,500 g/mol.

2.6.2 Molecular Weight Determination

Molecular weight estimation of heavy oils, residues, and asphaltenes is a very controversial area. Techniques that have been used to calculate the molecular weights of heavy oils include: vapor pressure osmometry, mass spectroscopy, florescence depolarization, and chromatography. Critics of VPO state that it measures aggregate molecular weight. Conversely, fluorescence depolarization critics state that it only detects molecules that fluoresce. Likewise, laser desorption mass spectrometry critics state that it induces fragmentation and measures the molecular weight of the smaller fragments. By using vapor pressure osmometry (with both toluene and 1,2 dichlorobenzene), Yarranton *et al.* concluded that the average molecule weight of a single asphaltene molecule is approximately 1,800 g/mol.[25] It was determined that average aggregates have molecular weights between 4,000 and 10,000 g/mol. The average weight of a single molecule was determined by extrapolating a plot of measured molecular weight versus asphaltene concentration to the intercept at zero asphaltene concentration. As the asphaltene concentration increases from 0 kg/m^3 to 20 kg/m^3, the measured molecular weight of the aggregates rapidly increases from about 1,800 g/mol to about 5,000 g/mol for both Athabasca and Cold Lake derived asphaltenes. Yarrangton *et al.* report that some of the variations in the reported molecular weight of asphaltenes might be due to different concentrations used in VPO experiments.

Florescence depolarization laser desorption mass spectrometry generally calculates lower molecular weights relative to vapr pressure osmometry. Based on florescence depolarization results, Mullins *et al.* determined that asphaltenes have an average molecular weight between 500 and 1,000 g/mol.[26] Asphaltene concentrations of 0.01 g/L were used in this study. Miller *et al.* used laser desorption mass spectrometry to calculate the average molecular weight of a Mayan asphaltene.[27] An average molecular weight of about 400 g/mol was calculated, with various asphaltene fractions having average molecular weights of between 300 and 600 g/mol.

2.6.3. NMR Spectroscopy

Nuclear magnetic spectroscopy (NMR) has given heavy oil researchers considerable insight into the chemical structure of heavy oil. The advantage of this spectroscopic technique is that it can be quantitative in addition to being analytical; it is capable of detecting a structural group and measuring its concentration. NMR spectroscopy is an example of a sophisticated technique that heavy oil researchers truly understand at a basic scientific fundamental level. As a result, heavy oil researchers have used NMR spectroscopy to gain new insights into the chemical structure of heavy oil. NMR spectroscopy gives the most accurate measurement of aromaticity - a very critical structural measurement.

The NMR spectra of heavy oils are rather noisy and difficult to interpret. Heavy oil researchers have used model compounds to resolve the various peaks found in both ^{1}H and ^{13}C-NMR spectra.[28] NMR spectroscopy is capable of resolving peaks associated with various carbon types: aromatic CH, aromatic C, linear aliphatic CH_2, alpha to aromatic, branched aliphatic CH, naphthenic CH, naphthenic CH_2, and CH_3.[28,29] Over time, new NMR spectrometry techniques, such as distortionless enhancement by polarization transfer (DEPT), are developed. As a result, more peaks are resolved and identified and new insights

into heavy oil structure occur. For example, implementation of DEPT, allowed the concentrations of C, CH, CH_2, and CH_3 groups to be calculated using ^{13}C NMR spectrometry.[30] Thus, by enhancing ^{13}C-NMR spectrometry with DEPT, significantly further insight into the molecular structure of heavy oils was made possible.

2.7. Asphaltenes

Asphaltenes represent the heaviest fraction of heavy oil and have a very large economic impact. By definition, asphaltenes are the component of oil insoluble in n-heptane and soluble in toluene. This solubility definition generally refers to the heaviest portion of crude oils with the highest aromatic contents. Asphaltenes can exist in high concentrations in heavy oil: Athabasca bitumen contains between 15-20% asphaltenes. Thus, their high weight compositions alone give them large significance in the upgrading, refining and processing of heavy oil. Asphaltenes are associated with many difficulties that occur during refining and processing. The heavy metal content of asphaltenes poisons catalysts. Asphaltene precipitation and fouling complicates transportation pipelines and makes refining more challenging. Conversely, high asphaltene content is desirable for paving and coating materials.

2.7.1. Two Alternative Asphaltene Models: Pericondensed and Archipelago

Asphaltenes contain C, H, N, Ni, O, S, and V and contain various types of paraffinic, naphthenic, and aromatic groups.[31] There is considerable debate regarding aspects of the structure of asphaltenes, especially with regard to the size of the aromatic groups and how they are linked to the other structural groups.[31-36] Two fundamentally different views of asphaltene structure are found in the current literature, where the asphaltene models can be categorized as

belonging to either the pericondensed or the archipelago structures. Pericondensed models are based on a core aromatic group containing a high number of fused rings (comprised usually of more than seven rings) with pendant aliphatic groups (Figure 2-8). The archipelago models suggest that asphaltenes consist of smaller aromatic groups linked by aliphatic bridges (Figure 2-9).

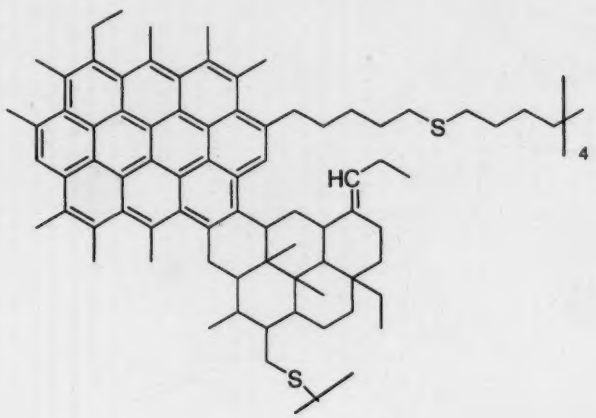


Figure 2-8. Pericondensed aspahltene representation of Yen *et al.*[35]

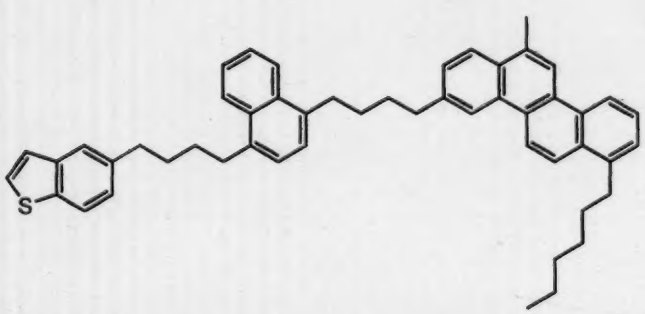


Figure 2-9. A general archipelago heavy oil structure.

2.7.1.1. The Pericondensed Model

2.7.1.1.1. X-ray Diffraction

X-ray diffraction was used by Yen *et al.* to gain insights into the chemical structure of asphaltenes.[35-37] Although Yen *et al.* correctly proposed that asphaltenes are composed of alternating series of *n*-alkyls and aromatic clusters, they significantly overestimated the degree of aromatic condensation. Yen *et al.* proposed asphaltene models that are currently known as pericondensed structures – a central highly condensed aromatic cluster with aliphatic side chains. Strausz *et al.* state that Yen *et al.* were later found to have significantly overestimated the x-ray diffraction signal attributed to the aromatic discs, thus overestimating the degree of condensation.[36] Strausz *et al.* also point out that, at the time of the Yen *et al.* studies, instrumental methods were unable to detect the presence of naphthenics and that few chemical approaches to structural determination existed.[36] Thus, the conclusions that were proposed based upon the X-ray work of Yen *et al.* illustrate that in structural chemistry it is never wise to propose definite conclusions based upon a single experimental technique. The results from multiple studies using different techniques and methodologies must be integrated to form conclusions.

2.7.1.1.2. Fluorescence Depolarization and Mass Spectroscopy

Mullins *et al.* used fluorescence depolarization to study asphaltenes.[26,38-40] Fluorescence depolarization analyses the anisotropic decay of the fluorescence emission of molecules in solution. A geometrical shape is first be assumed: either spherical or ellipsoid. Correlations - for given geometries - relate the rate of fluorescence decay to rotational correlation time. Once the rotational correlation time of a molecule is known, the radius of the molecule is calculated. Based upon fluorescence depolarization results, Mullins *et al.* concluded that asphaltenes contain one or two chromophores per molecule and have an

average molecular weight between 500 and 1,000 g/mol.[26] The major implication of small molecular weight asphaltenes that contain very few highly condensed aromatic groups is that asphaltene molecules will contain a very limited range of bonded attachments including thiophenes, sulfides, alkane chains, porphyrins, etc. Thus, if the pericondesned model of asphaltene structure is correct, asphaltenes are very chemically diverse. A molecule proposed by Groenzin and Mullins based upon fluorescence depolarization analysis is presented in Figure 2-10.[39]

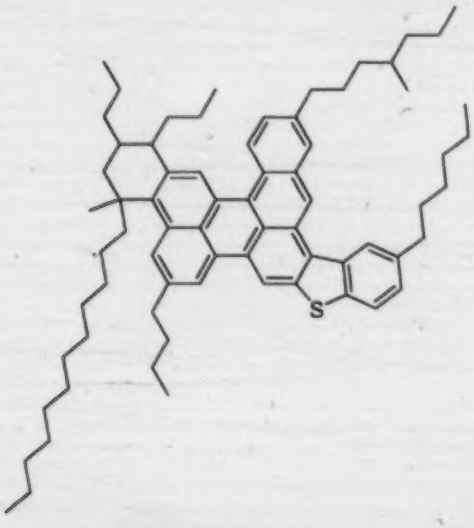


Figure 2-10. Asphaltene molecule proposed by Groenzin and Mullins.[39]

2.7.1.1.3. Criticism of Fluorescence Depolarization and Mass Spectroscopy

Strausz *et al.* point out a few concerns of using mass spectroscopy for calculating the molecular weight of asphaltenes.[41] According to Strausz *et al.*, it is challenging to select experimental conditions and instrumental parameters that will liberate single molecules from the molecular aggregate without cleavage of bonds. Thus, fragmentation is an issue associated with mass spectroscopy – the measured molecular weights might be the molecular weights of fragmentation

products. Strausz *et al.* also point out that it is challenging to insure that the vapor composition that enters the detection chamber is representative of the entire bulk sample.

Strausz *et al.* point out several problems associated with florescence depolarization technique used by Mullins *et al.* to calculate the molecular weights of asphaltenes.[41] These problems include:

1. Not all molecular species in asphaltenes necessarily absorb in the 300-700 nm range used by Mullins *et al.*
2. Not all absorbing molecules necessarily fluoresce. Florescence can be quenched by heavy atoms including S, intramolecular H-bonded complexes, and metal salts.
3. Aromatics in bridged structures exhibit free rotation about C-C and C-S bonds. As a result, the overall geometry of a molecule can not be approximated by a simple geometrical object. The calculated rotations in the florescence depolarization studies would be for the rotations rotations of the individual aromatic groups rather than the rotations of individual molecules.

Morgan *et al.* evaluated the effectiveness of both UV-fluorescence and UV-absorption as methods of detecting high molecular weight petroleum derived materials.[42] Maya asphaltenes, Maya heptane insolubles, and vacuum bottoms heptane insolubles from Chile were analyzed in this study. Each sample was characterized by two peaks: an early-eluting peak that corresponded to high molecular weight material excluded from the pores and a retained peak corresponding to relatively low molecular weight material that migrated through the pores. Compared to UF-A, UF-F was found to be significantly less sensitive at detecting material in the shorter elution times. Consequently, the number and average molecular weight values calculated using UV-A data were significantly higher than those calculated using UV-F data. For a vacuum bottoms heptane

insoluble fraction, the average molecular weight was calculated to be 1,608 g/mol using UV-A and 938 g/mol using UV-F. It was concluded that UV-F spectroscopy was not capable of detecting the full breadth of the molecular mass distribution in high molecular weight petroleum samples.

Gray also points out that the low molecular weight structures proposed by Mullins *et. al.* are not consistent with thermal cracking data.[43] When asphaltenes are cracked, a full range of products is observed: naphtha, light gas oil, gas oil, and vacuum residue. When the structures proposed by Mullins *et. al.* are cracked, naphtha and vacuum residue would dominate, the light gas oil and gas oil fractions would not result in significant amounts.

2.7.1.3. The Archipelago Model

The archipelago model is supported by data from VPO, pyrolysis, oxidation, thermal degradation and small angle neutron scattering (SANS) techniques.[25,31,34,36,44,45] The only structures that are consistent with both the molecular weights calculated using VPO and the aromacities determined using ^{13}C-NMR spectroscopy are archipelago structures. Mild pyrolysis of Athabasca asphaltenes produced an aromatic fraction containing alkylaromatics with 1-3 rings[34]. Highly condensed aromatic groups would not fragment into smaller aromatic groups during mild pyrolysis. Selective oxidation of Athabasca asphaltenes and interpretation of the relative distribution of reaction products also suggests that asphaltenes contain primarily small two to four ring aromatic groups and a small concentration of large, pericondensed aromatic groups.[34,36,44] The selective oxidation results also indicate that the aromatic clusters are linked by aliphatic chains up to 24 carbons long[36], thioether linkages[46], as well as biphenyl bridges.[36] Ni_2B reduction resulted in a 40% desulfurization of Athabasca asphaltenes and a fourfold molecular weight reduction.[46] The structure of high molecular weight asphaltenes is proposed to follow: [core]-S-[core]-S-[core]-S-[core][46]. Analysis of fractal aggregate dimensions of asphaltenes, obtained from

SANS analysis, indicated that the archipelago model is consistent with the analytical data.[47] When asphaltenes are thermally cracked, a full range of products arise in the naphtha (C_5-177°C), gas oil (343-524°C), and residue (524°C +) fractions.[43] Cracking of pericondensed asphaltenes such as those proposed by Groenzin and Mullins would yield products in only the gas, naphtha, and residue fractions.[43] In contrast, cracking of asphaltene molecules with archipelago structures would yield products with a wide range of molecular weights.

2.8. Hydrocarbon Thermal Cracking and Free Radical Mechanisms

2.8.1. Chain Reactions

Hydrocarbon thermal cracking involves free radical mechanisms.[48-50] A small number of reversible reaction steps both generate and consume free radicals. The reaction families involved in hydrocarbon pyrolyis include: homolytic dissociation, radical recombination, beta-scission, isomerization, and hydrogen abstraction. Each of these reactions is reversible. The families of the reverse reactions are: radical recombination, molecular disproportionation, and radical addition. Isomerizations and hydrogen abstraction reactions are reversible, but the reverse reactions are also denoted as isomerizations and hydrogen abstraction reactions. The reactions involved in hydrocarbon thermal cracking are presented in Figures 2-11 and 2-12. Initiation creates free radicals, propagation transfers radicals, and termination eliminates free radicals.

Initiation:

$M\text{-}V \rightarrow M^{\bullet} + V^{\bullet}$

Propagation:

Hydrogen abstraction: $R^{\bullet} + M \rightarrow RH + M^{\bullet}$
Beta- scission: $M\text{-}C\text{-}C\text{-}B^{\bullet} \rightarrow M\text{-}C{=}C + B^{\bullet}$

Termination:

radical + radical $\rightarrow$ products

Figure 2-11. Free reaction chain mechanism.[49]

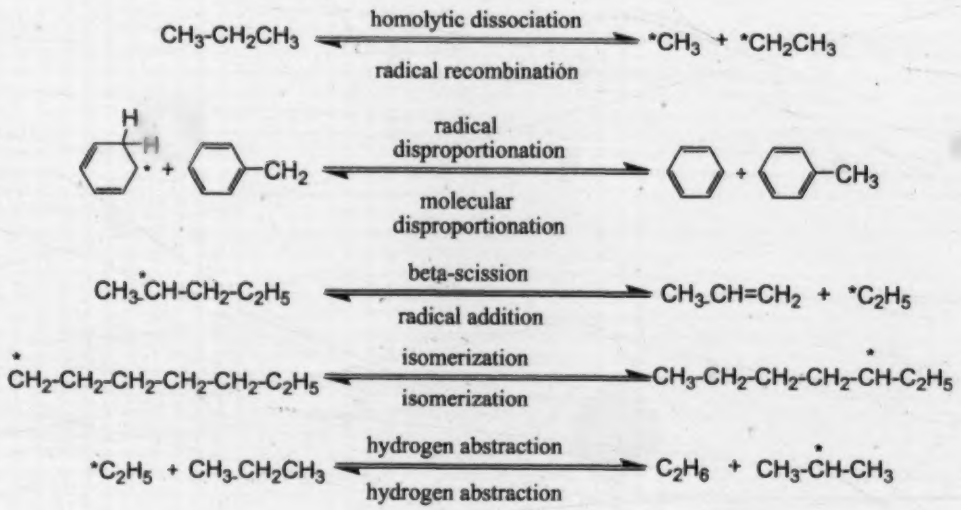


Figure 2-12. Free reaction chain mechanisms involved in hydrocarbon pyrolysis.[49]

High molecular weight hydrocarbons contain many atoms and many different C-C bonds. Hence, they contain many different potential reaction paths. When multiple paths exist, the fastest path is expected to be the path that contains the weakest bond - the bond with the lowest bond dissociation energy (Table 2-4). In general, C-C are generally weaker than C-H bonds, while C-C bonds are general stronger than C-S bonds. Primary bonds - both C-C and C-H - are stronger for primary carbons compared to secondary carbons. Double bonds or the presence of aromatic rings affect the strengths of both C-H and C-C bonds. The bonds in aromatics are unbreakable at normal process temperatures (<600 °C).

The refinery processes that crack residues into high value distillate fractions operate at 410-550 °C.[50] The cracking process is not just a bond breaking process. Rather than being a single step, the cracking process is composed of multiple steps that generate multiple radicals. The activation energy for breaking a C-C bond is very high (Table 2-4). However, Freund and Olmstead determined that the activation energy for cracking Arab Heavy and Cold Lake residues is approximately 213 to 217 kJ/mol lower than any of the bond dissociation energies in Table 2-4. The logic for this observation is that the propagation and termination steps have considerably lower activation energies than the initial initiation step. Thus, the overall reaction system will require relatively less energy per mol of residue, relative to the same amount of energy necessary to break up the residue via only the initiation reaction alone. As a result, the overall cracking reaction becomes more energetically favourable and proceeds significantly faster compared to the case where a single reaction was responsible for the bond breakages that occur during thermal cracking reactions.

Table 2-4. Bond Dissociation Energies.[49]

Chemical Bond	Representative bond	Bond energy (kJ/mol)
C-C (aliphatic)	C_2H_5-C_2H_5	365±4
C-H (primary)	H-$CH_2C_2H_5$	420 ±3
C-H (secondary)	H-$CH(CH_3)C_2H_5$	411±2.
C-C	CH_3-$CH_2C(CH_3)$=CH_2	301±3
C-C	CH_3-$CH_2C_6H_5$	317±6
C-S	CH_3S-C_2H_5	307±8
C-N	C_2H_5-NH_2	342±8
C-O	C_2H_5O-C_2H_5	344±4

2.8.2. Free Radical Reactions In Petroleum Cracking

Free radical chain reactions have been determined to be responsible for the cracking of alkanes, alkylaromatics, and cyclic saturated compounds.[48-50] Blanchard and Gray demonstrated the importance of free radical reactions in the cracking of vacuum residue[51]. 1-methyl naphthalene was added to Athabasca residue before it was thermally cracked. Radicals that were generated from the residue cracking would extract hydrogen from 1-methyl naphthalene and form stable benzylic radicals. The formation of stable benzylic radicals would decrease the concentration of radicals that participated in the chain reaction cracking mechanism. Thus, the rate of residue cracking would be expected to decrease. In this study, the residue conversion decreased from 48.6% to 35.7% when 1-methyl naphthalene was included. Likewise, a 3.64-fold increase of di-naphthyl species (a 1-methyl naphthalene termination product) was observed.

2.9. Molecular Kinetic Simulation of Thermal Cracking

Mass lumped kinetics represents the easiest method to model the cracking of a heavy oil. In a mass lumped kinetic model, a heavy oil is represented as a series of distillation fractions. Kinetics is then used to model the mass conversion of each lump into each other lump. The advantage of mass lumped kinetic models is that they are easy to understand, it is easy to obtain distillation experimental data to construct them, and they are mathematically straightforward. Unfortunately, mass lumped kinetic models are generally only valid for specific feeds, catalysts and sets of process conditions.

Molecular based kinetic models are more promising that mass lumped kinetic models. The major disadvantage of molecular based kinetic models is that they require detailed information about the molecules that compose a reacting feed fraction. Thus, ^{13}C-NMR spectroscopy, ^{1}H-NMR spectroscopy, elemental analysis and other analytical data are required. It is often costly and time intensive to obtain these data. One also must construct models of the feed to use molecular based kinetic models. This creates a complex situation. If the original feed model was not correct, it would then be impossible to generate a correct reaction model. However, there are several advantages associated with molecular kinetic models. Molecular kinetic models describe how individual molecules react at the molecular level. Thus, these models can describe how molecules react under different temperatures, pressures, and other reaction conditions. A molecular kinetic model can describe the reactivity of any feed as long as it can be characterized analytically. Thus, although there are complications associated with molecular based kinetic models, they have the potential to be very robust and versatile models that can describe the cracking of a variety of different feeds under a variety of different processing conditions.

2.9.1. Froment's Molecular Kinetic Thermal Cracking Simulations

Froment created a mechanistically detailed reaction technique to model thermal cracking reaction processes of simple hydrocarbon molecules.[52] This approach models several fundamentals associated with the cracking of hydrocarbons, including: initiation, radical isomerisation, radical decomposition, hydrogen abstraction. In the case of 3-methylheptane, 35 different types of reactions are involved in the reaction scheme. For larger and more complicated molecules, it would be very challenging and time intensive to create a reaction scheme. The approach that Froment developed, first characterizes each feed molecule using a binary matrix that documents all C-C bonds in a molecule. Each type of reaction is modeled as a matrix operation. When the matrix operation for a specific type of reaction is applied to the feed molecules, matrices that correspond to product molecules are created. Thus, matrix operations are sequentially applied to generate all potential products. The number of differential equations that must be solved, and the number of variables involved, are both reduced by expressing certain radical concentrations in terms of other intermediate products. Thermodynamics is used to calculate some reaction rate parameters for reversible reactions including: initiation, termination, decomposition, and addition.[53,54] In the case of naphtha cracking, Willems and Froment were able to reduce the number of rate parameters from 140 to 60.[53]

Froment successfully applied his reaction modeling methodology to model the reactivity of a variety of different processes including: cracking of vacuum gas oil, catalytic methanol to olefin process, and the cracking of naphthas.[55,56] Froment states that the cracking network of n-dodecane contains 201 n- and i-paraffins and 1,123 olefins, while the cracking network of n-pentylcyclohexane contains 2,697 naphthenic and acyclic components.[5] Thus, even relatively small molecules are associated with very large reaction networks. The rate coefficients depend both on the structure of the reacting group and the immediate

environment of the reacting group. Transition state chemistry and statistical thermodynamics are used to calculate rate constants. Gaussian distribution modeling was used by Froment to model the feed distribution of a gas oil.[55] A single Gaussian distribution was used to model the mole fraction distribution of various hydrocarbons in a gas oil including: single ring aromatics, di-ring aromatics, tri-ring aromatics, mono-ring napthenes, di-ring napthenes, etc.

2.9.2. Dente's Molecular Kinetic Thermal Cracking Simulations

Dente *et al.* developed detailed mechanistic models of cracking for the simulation of cracking reactions of light gases , naphthas, and gas oils[57,58]. The kinetic algorithm used by Dente *et al.* handled 2000 different reactions and 86 different molecular species. Rate constants were estimated using experimental data and often approximated to be very low. Certain isomer radical intermediates were lumped together to simplify the simulations. The limited computational power at the time required that simplifications had to be made. Material, energy, and mechanical balances were also simultaneously solved. The pyrolysis kinetic simulator of Dente et al. was capable of accurately predicting the performance of steam reactor coils and was useful in the design, optimization, and control of furnace coils.

2.9.3. Joshi's Molecular Kinetic Thermal Cracking Simulations

Joshi *et al.* used an approach that is very similar to Froment's and Dente's approaches to model the cracking of a gas oil.[59] A major difference between Joshi's approach compared to the other two approaches, is that Joshi *et al.* used a molecular representation to model the feedstock. The input to the cracking model that Joshi *et al.* developed is a molecular representation created using Campbell and Klein's Monte Carlo algorithm.[10] Just like the approach that

Froment used, Joshi *et al.* represented the reactions associated with gas oil cracking as a series of elementary steps: protolytic cleavage, beta-scission, hydride/methyl shift, protonation/deprotonation, and PCP isomerisation. Each type of reaction had an associated matrix operation that will identify reaction products and create matrices for them. Each type of reaction has a detailed series of rules that specify the type of chemical groups that can and cannot participate in them. The rate constants for all reactions were described using an Arrhenius A factor and a QSRC formula that relates the activation energy to the enthalpy change of a reaction (Equations 2-1 and 2-2).

$$E_A = E_0 + \alpha^* \Delta H_{rxn} \qquad (2\text{-}1)$$

$$k_{ij} = A_j^* \exp(-(E_{0,j} + \alpha_j \Delta H_{rxn,i}))/RT) \qquad (2\text{-}2)$$

The Arrhenius A values were found from the literature and alpha values were set to 0.5. The values of the activation energies were found by mathematical optimization and the heats of reaction were calculated using MOPAC. The model was first built up using the matrix operations and their associated rules. The gas oil representation contained 194 feed molecules. Two-hundred and thirty-one catalyst surface species were also included in the representation. In total, 5,500 reactions were required. Pilot plant data was used to both build up and test the model. In total, nine activation energies were optimized. A catalyst dependent parameter, a coking rate constant, and a deactivation parameter were also optimized. The predicted product distribution of the Joshi *et al.* cracking model closely matched the pilot plant data.

2.9.4. Monte Carlo Thermal Cracking Simulation

Neurock and Klein used a Monte Carlo simulation to model Asphaltene hydroprocessing.[8] A molecular representation containing 10,000 molecules was first created for the asphaltene fraction. Alkyl side chain thermolysis, aromatic hydrogenation and saturated ring dehydrogenation were modeled in the cracking simulations. Model compounds were used to calculate rate constants for the

various reactions. During the reaction simulations, reaction sites were identified and relative probabilities of reaction were determined based on the reaction rate constants derived from the model compounds. Using first order kinetics, the time for each reaction step was determined. A simulation proceeded until a final predetermined final time was reached.

2.10. Group Contribution Methods

Reaction simulations require physical property data. Boiling points, kinetic rate constants, and critical properties are usually required by reaction simulations. For homogenous mixtures, the properties that are required can often be measured in a laboratory. However, it is often impractical, cost prohibitive, or not feasible to calculate the physical properties of a large variety of molecules. Residue petroleum fractions represent an unique case. It is not possible to isolate individual molecules in residue fractions. As a result, it is not possible to experimentally determine the physical properties of individual molecules in petroleum fractions. However, physical properties of individual molecules are necessary for reactivity and other simulations. Group contribution is a method that allows various physical properties to be calculated for reactivity simulations.

Group contribution approaches allow physical properties to be calculated from petroleum molecular representations. Traditional mass lumped kinetics can predict the mass transformations that result when petroleum is processed. However, mass lumped kinetics does not provide a framework to predict the physical properties that result when petroleum is processed. It is relatively straightforward to calculate physical properties of molecular representations using group contribution methods. Molecular representations can therefore be used to react in molecular simulations, and to calculate the physical properties of the resulting product molecules. Thus, group contribution methods have a key

role in making molecular representations practical for engineers to design effective processing models. Group contribution methods are beginning to have a significant role in creating molecular representations. Experimental methods can be used to calculate the physical properties of petroleum fractions. When a molecular representation is created for a petroleum fraction, a group contribution method can be used to calculate the physical properties of individual molecules in the petroleum fraction. Averaging methods can then calculate the physical properties of an entire petroleum fraction using the physical properties of the individual molecules. Thus, group contribution methods allow experimental physical property data to be used when molecular representations are being created. Group contribution methods will therefore have very important roles in both the creation of molecular representations and in the development of molecular simulations.

Group contribution methods are fundamentally straightforward. A statistically large number of molecules with experimentally determined physical properties are studied. The molecules are described using a relatively small number of descriptive groups. A statistical method is then used to determine the incremental contribution of each group to each physical property.

2.10.1. Joback and Reid Group Contribution Method

The first widely used group contribution method was proposed by Joback and Reid.[60] The Joback and Reid group contribution method allowed the following critical properties to be estimated from the structure of a molecule: normal boiling point, normal freezing point, critical pressure, critical temperature, critical volume, enthalpy of formation, Gibbs energy of formation, heat capacity, enthalpy of vaporization, enthalpy of fusion, and liquid viscosity. For each of these properties, estimation equations were proposed. The variable in each equation, was the summation of the contribution of each descriptive group. In total, 41 molecular groups were used to describe molecules. These groups

consisted of groups that described "nonring" carbon, "ring" carbon, halogen increments, oxygen increments, nitrogen increments, and sulphur increments. Nonlinear regression was used to calculate the contribution. Each of the 41 groups to each physical property-using the proposed estimation equation for each variable. For each physical property, between 155 and 438 molecules were used. For boiling point, an average percent error of 3.6% was reported. Conversely, the error for critical pressure was reported as being 5.2%. The Joback and Reid method has not stood the test of time. The defining groups in this method are not specific enough - too few molecules were used when this group contribution was developed. In addition, relatively small molecules with relatively low aromatic condensation levels were used during its development. As a result of these drawbacks, this group contribution method is not applicable to high molecular weight petroleum fractions. However, this method was a historical precursor, and allowed the development, of many other group contribution methods.

2.10.2. Marrero-Gani Group Contribution Method

The Marrero-Gani group contribution method is applicable to relatively large petroleum molecules.[61] Compared to the Joback-Reid group contribution method, the Marrero-Gani is very precise and very explicit. More than 2000 compounds containing between three and sixty carbon atoms were used to develop this method. Some large and polycyclic compounds were used is the development of this contribution method. The molecules were described by 182 first-order groups, 122 second-order groups, and 66 third-order groups. Regression was used to determine the contribution of each group to each physical property. The first order groups are as small as possible and describe a wide variety of organic compounds: aromatics, naphthenics, alcohols, phenols, ketones, acids, esters, etc. First order groups describe an entire molecule. The second-order groups describe polyfunctional compunds and allow differentiation among isomers while the third-order groups describe systems of fused and

unfused rings systems. Overlapping of the different orders of groups is allowed. The Marrero-Gani group contribution method is capable of calculating the following physical properties: normal boiling point, critical temperature, critical temperature, critical pressure, critical volume, standard enthalpy of formation, standard Gibbs energy, normal melting point and standard enthalpy of fusion. The explicit nature of this group contribution method allows it to be used to calculate the physical properties of a wide range of molecules.

2.11. Petroleum Model Compounds

The molecular weights and structural complexities of asphaltenes and petroleum residues currently render it impossible to create adequate model compounds for them. Ideally, one could synthesize model compounds that are approximately the same size, have the same functional groups, and have the same overall topology and organization as asphaltenes and residues. If this was possible, asphaltene scientists and engineers would have ideal compounds to gain insights into the structural chemistry, reactivity, and phase behavior of high molecular weight petroleum molecules.

2.11.1. Model Compound Thermal Cracking Studies

Over the past twenty years, small model compounds have been used to study how various functional groups influence the reactivity of high molecular weight petroleum molecules. The logic behind these reactivity studies is straightforward, reactivity and processability is independent of molecular weight since reactions are the result of the breakage and/or rearrangement of chemical bonds. A functional group will react exactly the same in a small molecule, under the same reaction conditions, compared to a larger molecule that contains the same functional group. Thus, the insights and knowledge acquired by studying small molecules that contain the same functional groups as high molecular

weight petroleum is directly applicable to high molecular weight petroleum. Therefore model compound reaction studies are important to generate data for reaction simulations.

However, one must be very careful when utilizing small compound data to gain insights into how larger petroleum molecules react. Some small model compound molecules are very difficult to perform experiments with and analyze due to their physical properties such as volatility. One must look at the mass balance associated with any reactivity study that utilizes small compounds. When a product mass balance deviates significantly from 100%, one can not use measurements from the study for further calculations or computations. Thus, although these studies give insights into how a molecule reacts, the concentration and rate values published in these studies should not be used to calculate reaction rates or distribution of reaction products of higher molecular weight petroleum molecules.

The thermal cracking of a variety of small model compounds, representative of various functional groups that are found in high molecular weight petroleum molecules, has been investigated by many different researchers. The model compounds presented in Figures 2-13 to 2-17 were synthesized to study the thermal cracking of the following functional groups: phenyl, large aromatics, naphthenic rings, thioethers, branched carbons, aromatic side chains, and long carbon side chains. The thermal reaction studies are discussed in greater detail – along with product distribution data – in Appendix A. In these reactivity studies, the model compounds were thermally cracked at temperatures ranging between 400 and 500℃. The reaction products were separated, identified, and quantified using gas chromatography. An example of the type of data presented in one of these model compound reactivity studies is presented in Table 2-5.

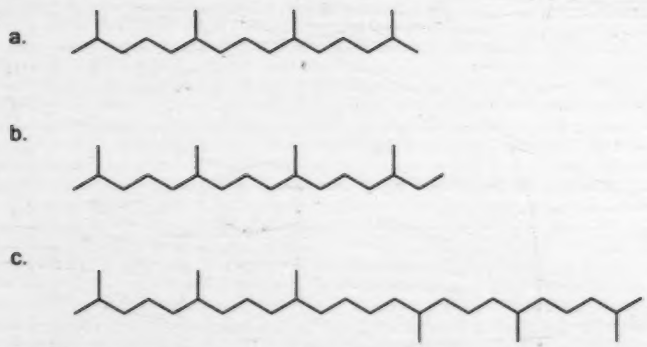


Figure 2-13. The structures of pristane (a.), phytane (b.), and squalane (c.).[62]

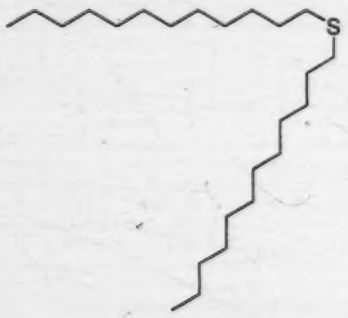


Figure 2-14. The structures of didodecyl sulfide.[63]

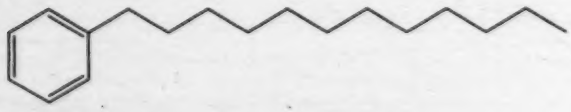


Figure 2-15. The structure of phenyldodecane.[64]

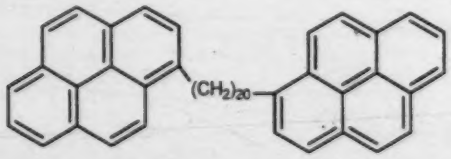


Figure 2-16. The structure of 1,20-di(1-pyrenyl)eicosane.[65]

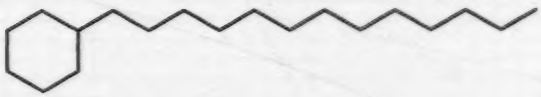


Figure 2-17. The Structures of: a. Tridecylcyclohexane.[66]

Table 2-5. Molar yields of products from tridecylcyclohexane thermolysis at 400 ℃.[66]

Time (min)	30	90	180
cyclohexane	0.59	1.58	3.14
cyclohexene	0.27	0.73	1.4
heptene	0.09	0.43	0.68
heptane	0.1	0.54	1.26
methylcyclohexane	0.22	0.65	1.49
methylenecyclohexane	0.63	1.01	1.96
toluene			
methylcyclohexenes	0.1	0.81	1.1
octene	0.06	0.17	0.4
octane	0.13	0.43	1.01
cyclohexylethene	0.17	0.39	0.61
ethylcyclohexane	0.18	0.54	1.17
nonene	0.06	0.25	0.44
nonane	0.07	0.39	0.94
cyclohexylpropene	0.15	0.36	0.51
propylcyclohexane	0.14	0.42	0.98
decene	0.05	0.18	0.32
decane	0.07	0.4	0.93
cyclohexylbutene	0.12	0.27	0.33
butylcyclohexane	0.11	0.35	0.75
undecene	0.05	0.17	0.27
undecane	0.07	0.39	0.87
cyclohexylpentene	0.1	0.24	0.28
pentylcyclohexane	0.13	0.38	0.82
dodecene	0.06	0.2	0.27
dodecane	0.55	1.46	2.91
cyclohexylhexene	0.06	0.23	0.26
hexylcyclohexane	0.13	0.37	0.76
tridecene	0.34	0.66	0.75
tridecane	0.19	0.55	1.12
cyclohexylheptene	0.12	0.22	0.24
heptylcyclohexane	0.17	0.42	0.83
cyclohexyloctene	0.05	0.17	0.21
octylcyclohexane	0.11	0.33	0.67
cyclohexylnonene	0.05	0.15	0.14
nonylcyclohexane	0.1	0.3	0.58
cyclohexyldecene	0.05	0.13	0.13
decylcyclohexane	0.26	0.46	0.68
tridecylcyclohexane	87.4	88.8	61.1
alkanes	1.18	4.15	9.05
olefins	0.75	2.06	3.14
alkylcycolohexanes	2.15	5.8	11.9
cyclohexylalkenes	1.77	3.9	6.07

2.11.2. Asphaltene Model Compounds

The model compounds used in the reaction studies described in section 2.11.1 have relatively low molecular weights and although they are representative of specific chemical groups found in heavy oil molecules, they are not representative of individual heavy oil molecules. When researchers have model compounds that are truly representative of the molecules present in bitumen, several possibilities will arise. With model heavy oil compounds it will be possible to fine tune and improve the NMR spectroscopic analysis of bitumens. Likewise, the molecular representation methodology will be improved and refined by modeling the composition of mixtures of known compositions. Furthermore, reactivity studies with the model heavy oil compounds will result in more accurate thermal reactivity simulations being developed.

Organic synthetic chemistry has recently been used to synthesize a series of model heavy oil compounds: 4,4'-bis-(2-pyren-1-yl-ethyl)-[2,2']bipyridinyl (PBP), C6-Hexabenzocoronene, and C9- Hexabenzocoronene (Figure 2-18).[67,68] Vapor pressure osmometry, NMR spectroscopy, calorimitry, and solubility all provide evidence that these model compounds self associate in solution. These model compounds do not fully represent the chemical richness and diversity detected in bitumen samples. However, these compounds represent very promising starting point to synthesize additional heavy oil model compounds in the future.

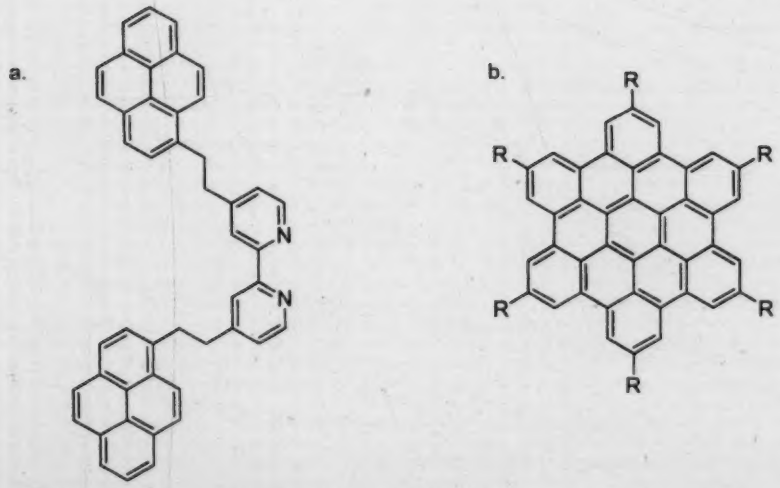


Figure 2-18. The structures of: a. 4,4'-bis-(2-pyren-1-yl-ethyl)-[2,2']bipyridinyl (PBP); b. Hexabenzocoronene (R= C_6 or C_9).[67,68]

2.12. References

1. Brown, J.K. Ladner, W.R. Sheppard, N. *Fuel,* **1960**, *39*, 79-86.
2. Brown, J.K. Ladner, W.R. Sheppard, N. *Fuel*, **1960**, *39*, 87-96.
3. Yen, T.F. *Prepr. Div. Petrol. Chem. Soc.* **1972**, *17(4)*, F102-F110.
4. Bartle, K.D. Martin, T.G. Williams, D.F. *Fuel.* **1975**, *54*, 226-235.
5. Oka, M. Chang, H.C. Gavalas, G. R. *Fuel*, **1977**, *56*, 3-8.
6. Quann, R.J.; Jaffe, S.S.; *Industrial and Engineering Chemical Research.* **1992**, *31*, 2483-2497.
7. Jaffe, S.B.; Freund, H.; Olmstead, W.N. *Industrial and Engineering Chemical Research*. **2005**, *44*, 9840-9852.
8. Neurock, M.; Libanati, C.; Klein, M.T. Modelling asphaltene reaction pathways: intrinsic chemistry. *AIChE Symposium Series.* **1989**, *273(85)*, 7-14.
9. Neurock, M.; Nigam, A.; Trauth, D.; Klein, M.T. *Chem. Eng. Sci.* **1994**, *49*, 4153-4177.
10. Campbell, D M.; Klein, M.T. *Applied Cat. A: General* **1997**, *160*, 41-54.
11. Khorasheh, F.; Khaledi, R.; Gray, M.R. *Fuel.* **1998**, *77 (4)*, 247-253.

12. Murgich, J.; Merino-Garcia, D.; Andersen, S.I.; Manuel del Romo , J.;Galeana, C.L.;
Langmuir. **2002**, *18*(23), 9080-9086.

13. Murgich, J. *Molecular Simulation.* **2003**, *29(6-7)*, 451-461.

14. Murgich, J.; Abanero, J.A.; Strausz, O.P. *Energy & Fuels.* **1999**, *13(2)*, 278-286.

15. Murgich, J. *Petroleum Science and Technology.* **2002**, *20(9-10)*, 983-997.

16. Murgich, J.; Strausz, O.P. *Petroleum Science and Technology.* **2001**, *19(1-2)*, 231-243.

17. Mendeleyev, B.; Humboldt, A. von. *Voyage aux Régions Equinoxales du Nouveau Continent Pendant les Années 1799-1804.* **1877**, 229

18. Berthelot, M. Sur *L'origine des Carbures et des Combustibles Minéraux, Ann. Chim. Phys.* **1866**, *4e Ser. 9*, 481.

19. Engler, C. Zur Bildung des Erdöls, Chem Ber. **1888**, *21*, 1816.

20. Triebs, A. *Chlorophyll und Haminderivat in Bituminosen Gesteinen, Erdolen, Erdureschsen and Asphalten, Liebigs Ann. Chem.* **1934**, *510*, 42.

21. Gray, M.R. *Upgrading Petroleum Residues and Heavy Oils.* **1994**. New York: Marcel Dekker, Inc.

22. Speight, J.G. *The Chemistry and Technology of Petroleum.* **1999**. New York: Marcek Dekker, Inc.

23. Ignasiak, T.; Kemp-Jones, A.V.; Strausz, O.P. *Journal of Organic Chemistry*, **1977**, *42*, 312-320.

24. Chung, K.H.; Xu, C.M.; Hu, X.Y.; Wang, R.N. *Oil Gas J.* **1997**, *95(3)*, 66-70.

25. Yarrangton, H. W.; Alboudwarej, H.; Jakher, R. *Ind. Eng. Chem. Res.* **2000**, *39(8)*, 2916-2924.

26. Cunico, R.L.; Sheu, E.Y.; Mullins, O.C. *Petroleum Science and Technology.* **2004**, *22 (7-8)*, 787-798.

27. Miller, J.T.; Fisher, R.B.; Thiyagarajan, P.; Winans, R.E.; Hunt, J. E. *Energy & Fuels.* **1998**, *12(6)*, 1290-1298.

28. Thiel, J.; Gray, M.R. *AOSTRA J. Res.* **1988**, *4*, 63-73.

29. Snape, C.E.; Ladner, W.R.; Bartle, K.D. *Anal. Chem.* **1979**, *51(13)*, 2189-2198.

30. Kotlyar, L.S.; Morat, C.; Ripmeester, J.A. *Fuel* **1991**, *70(1)*, 90-94.

31. Strausz, O.P.; Peng, P. Murgich, J. *Energy Fuels* **2002**, *16*, 809-822.

32. Groenzin, H.; Mullins, O.C. *J. Phys. Chem A* **1999**, *103*, 11237-11245.

33. Groenzin, H.; Mullins, O.C. *Energy Fuels* **2000**, *14*, 677-684.

34. Strausz, O.P.; Mojelsky, T.W.; Faraji, F.; Lown, E.M. *Energy Fuels* **1999**, *13(2)*, 207-227.

35. Yen, T.F. *Prepr. Div. Petrol. Chem. Soc.* **1972**, *17(4)*, F102-F110.

36. Strausz, O.P.; Mojelsky, T. W.; Lown, E. M. *Fuel*, **1992**, *71*, 1355-1363.

37. Yen, T.F.; Dickie, J.P. *Journal of The Institute of Petroleum.* **1968**, *54*, 50-53.

38. Groenzin, H.; Mullins, O.C. *J. Phys. Chem A* **1999**, *103*, 11237-11245.

39. Groenzin, H.; Mullins, O.C. *Energy Fuels* **2000**, *14*, 677-684.

40. Ralston, C.Y.; Mitra-Kirtley, S.; Mullins, O.C. *Energy & Fuels.* **1996**, *10(3)*, 623-630.

41. Strausz, O.P.; Peng, P. Murgich, J. *Energy Fuels* **2002**, *16*, 809-822.

42. Morgan, T.J.; Behrouzi, M.; Herod, A.A.; Kandiyoti, R. *Energy and Fuels.* **2005**, *19(1)*, 164-169.

43. Gray, M. R. *Energy & Fuels.* **2003**, *17(6)*, 1566-1569.

44. Strausz, O. P.; Mojelsky, T. W.; Lown, E. M.; Kowalewski, I.; Behar, F.; *Energy Fuels* **1999**, *13(2)*, 228-247.

45. Gawrys, K.L.; Spiecker, P.M.; Kilpatrick, P.K. *Pet. Chem. Div. Pre* **2002**, *47(4)*, 332-335.

46. Peng, P.; Morales-Izquierdo, A.; Hogg, A.; Strauz, O.P. *Energy Fuels* **1997**, *11*, 1171-1187.

47. Gawrys, K.L.; Spiecker P.M.; Kilpatrick, P.K. *Pet. Sci. Tech.* **2003**, 21 (3-4): 461-489.

48. Poutsma, M.L. *J. Anal. Appl. Pyrolysis.* **2000**, *54*, 5-35.

49. Savage, P.E. *J. Anal. Appl. Pyrolysis.* **2000**, *54*, 109-126.

50. Gray, M. R.; McCaffrey, W. C. *Energy & Fuels.* **2002**, *16*, 756-766.

51. Blanchard, C.A.; Gray, M.R. *Abstr Pap Am Chem Soc.* **1997**, *42*, 137-141.

52. Clymans, P.J.; Froment, G.F. *Computers and Chemical Engineering.* **1984**, *8(2)*, 137-142.

53. Willems, P.A.; Froment, G.F. *Ind. Eng. Chem. Res.* **1988**, *27*, 1959-1966.

54. Willems, P.A.; Froment, G.F. *Ind. Eng. Chem. Res.* **1988**, *27*, 1966-1974.

55. Moustafa, T.M.; Froment, G.F. *Ind. Eng. Chem. Res.* **2003**, *42 (1)*, 14-25.

56. Alwahabi, S.M.; Froment, G.F. *Ind. Eng. Chem. Res.* **2004**, *43(17)*, 5098-5111.

57. Dente, M.; Ranzi, E.; Goossens, A.G. *Computers & Chem. Eng.* **1979**, *3*, 61-75.

58. Dente, M.; Pierucci, S.; Ranzi, E.; Bussani, G. *Chem. Eng. Sci.* **1992**, *47 (9-11)*, 2629-2634.

59. Joshi, P. Iyer, S. Klein, M. *Reviews in Process Chemistry and Engineering.* **1998**, *1*, 111-140.

60. Joback, K.G.; Reid, R.C. *Chemical Engineering Communication.* **1987**, *57*, 233-243.

61. Marrero, J.; Gani, R.; *Fluid Phase Equilibria*, **2001**, *183-184*, 183-208.

62. Kissin, Y.V. *Ind. Eng. Chem. Res.* **1987**, *26 (8)*, 1633-1638.

63. Abikhers, V.; Fixari, B.; Leperchex, P. *Fuel.* **1986**, *65 (3)*, 442-446.

64. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1987**, *26(2)*, 374-376.

65. Freund, H.; Matturro, M.G.; Olmstead, W.N.; Reynolds, R.P.; Upton, T. H. *Energy & Fuels.* **1991**, *5(6)*, 840-846.

66. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1988**, *27*, 1348-1356.

67 Tan XL, Fenniri H, Gray MR. *Energy & Fuels.* .**2008**, *22(2)*, 715-720.

68 Rakotondradany F, Fenniri H, Rahimi P, Gawrys, K.L., Kilpatrick, P. K., Gray, M.R.: *Energy & Fuels.* **2006**, 20(6), 2439-2447.

Chapter 3. Quantitative Molecular Representations of Athabasca Asphaltenes[1]

3.1. Background

Quantitative molecular representations integrate information from various analytical techniques to yield an overall picture of a petroleum sample. The average molecule approach of interpreting petroleum analytical data began in the early 1960s.[1,2] This approach uses correlations that rely on ^{1}H-NMR spectroscopy data to calculate average properties, including aromaticity and degree of substitution.[1,2] When many different molecules represent a petroleum fraction, their process chemistry and product properties can be modeled. More detailed representations have been created using Monte Carlo methods based on data from ^{1}H-NMR spectroscopy, elemental analysis, and VPO for various asphaltenes and residues.[3,4] The number of molecules contained in each of these representations ranges from 12 to 10,000 molecules. These methods have been extended to use structural data from quantitative ^{13}C-NMR spectroscopy.

3.2. Objective

The goal of this work was to generate a quantitative molecular representation of Athabasca asphaltene that was consistent with data from elemental analysis, ^{1}H and ^{13}C-NMR spectroscopy, and VPO. The method of construction followed the Monte Carlo approach of Klein and co-workers[3-5] but incorporated ^{13}C-NMR spectroscopy data, a distribution of aromatic cluster size following the archipelago framework[6,7,8,9,10], and both alkyl[8] and thioether bridges.[10] The representation method was designed to facilitate simulations of reactions and to allow representation of other residue fractions.

---

[1] 1. A version of this Chapter was published in Energy & Fuels, **2004**, *18*, 1377-1384.

3.3. Materials and Methods

3.3.1. Materials

A fraction of asphaltene material obtained from supercritical fluid extraction of Athabasca vacuum residue was modeled in this study.[11] Athabasca bitumen was first separated by vacuum distillation. The residue (524+ °C) was then separated into ten fractions using supercritical fluid extraction. This asphaltene rich fraction contained 88% by weight n-C_5 asphaltenes and the balance heavy maltenes. All of the asphaltenes were concentrated into the heaviest fraction, which accounted for approximately 40% of the vacuum residue.

3.3.2. Analytical Data

The NMR spectroscopic analyses and interpretations in this study were performed by Dr. Heather Dettman at NCUT. Quantitative ^{1}H- and ^{13}C-NMR spectroscopy analyses were performed using a Varian XL-300 NMR spectrometer operating at 299.943 MHz for proton and 75.429 MHz for carbon. Deuterocholoroform ($CDCl_3$) was used to dissolve the petroleum samples. The proton data were acquired using a 2.5 s acquisition time, a 6000 Hz sweepwidth, a 30.8° pulse flip angle, and no recycle delay. The proton spectra consisted of 128 scans and were referenced to the residual chloroform ($CHCl_3$) resonance at 7.24 ppm. A line broadening of 0.3 Hz was used for signal-to-noise improvement. The carbon spectra were collected with a 0.9 s acquisition time, a 16,500 Hz sweep width, a 32° pulse flip angle, and a recycle delay of 4 s. Reverse-gated waltz proton decoupling was also used. The carbon spectra resulted from 10,000 scans and were referenced to the $CDCl_3$ resonance at 77 ppm. 10-Hz line broadening was used for signal-to-noise improvement.

The assignments of regions in the NMR spectra were based on model compounds[12,13] and two dimensional NMR spectroscopy techniques including HETCOR[14] (heteronuclear shift correlation), COSY[14](homonuclear chemical shift correlation) and DEPT[15,16] (distortionless enhanced polarization transfer). The identified carbon types are represented in Figure 3-1. The assignments for ^{1}H-NMR spectroscopy are summarized in Table 3-1 and those for ^{13}C-NMR spectroscopy in Table 3-2. The NMR spectra associated with asphaltenes and residues are very complex and difficult to interpret. In some cases, NMR spectral regions were assigned to more than one type of proton or carbon. The quantities of hydrogen and carbon types listed in Tables 3-1 and 3-2 were calculated directly from the proton and carbon NMR spectral integrals. A number of chemical species were determined using assumptions or balances. A listing of these groups and how they were calculated is presented in Table 3-3. For example, to calculate aliphatic CH, the signal associated with region CP1 and half the signal associated with region CP2 were assigned to aliphatic CH groups. The concentrations of the chemical groups in terms of % C are presented in Table 3-4. The C/H ratio from the property concentrations agreed with the data from elemental analysis. Along with elemental analysis data and VPO results from Chung *et al.*[11], propagation of error formulae were used to estimate the error associated with the value of each group (Table 3-4).

Molecular weight was determined by VPO at 130 ℃ using *o*-dichlorobenzene as the solvent in order to minimize molecular associations. The average molecular weight of the asphaltene sample was used to express the carbon groups in terms of mol group/mol molecule in order to construct representative molecules.

Table 3-1. NMR spectroscopy chemical shifts of proton spectral regions.

Region	Chemical Shifts (ppm)	Structural type
HA1	10.7 to 7.4	Polyaromatic
HA2	7.4 to 6.2	Monoaromatic
HO1	6.2 to 5.1	Olefinic CH
HO2	5.1 to 4.8	Olefinic CH_2
HO3	4.8 to 4.3	Olefinic CH_2
HP1	4.3 to 2.4	α to Aromatic
HP2	2.4 to 2.0	α to Aromatic
HP3	2.0 to 1.09	Aliphatic CH_2
HP4	1.09 to –0.5	γ-CH_3

Table 3-2. NMR spectroscopy chemical shifts of carbon spectral regions.

Region	Chemical Shifts (ppm)	Structural Type
CA1	190 to 170	Carbonyl and acid carbon
CA2	170 to 129	Quaternary C (Q_1 + Q_2)
CA3	129 to 115.5	Aromatic CH (C_1 + C_2)
CA4	115.5 to 113.5	Olefinic CH_2
CA5	113.5 to 100	Olefinic CH_2
CP1	70 to 45	Aliphatic CH
CP2	45 to 32.7	Aliphatic CH and CH_2
CP3	32.7 to 30.8	Aliphatic γ-CH_2, β to aromatic CH_2,
CP4	30.8 to 28.5	Other aliphatic carbon (chain δ-CH_2, α-to aromatic naphthenes, aromatic-attached β-CH_2)
CP5	28.5 to 25.0	Naphthenic CH_2
CP6	25.0 to 21.9	"Other aliphatic" carbon (chain β-CH_2, α-to aromatic or isobutyl CH_3)
CP7	21.9 to 17.6	Ring α-CH_3
CP8	17.6 to 14.7	β-CH_3
CP9	14.7 to 12.3	Chain α-CH_3
CP10	12.3 to 0	Branched-aliphatic terminal CH_3

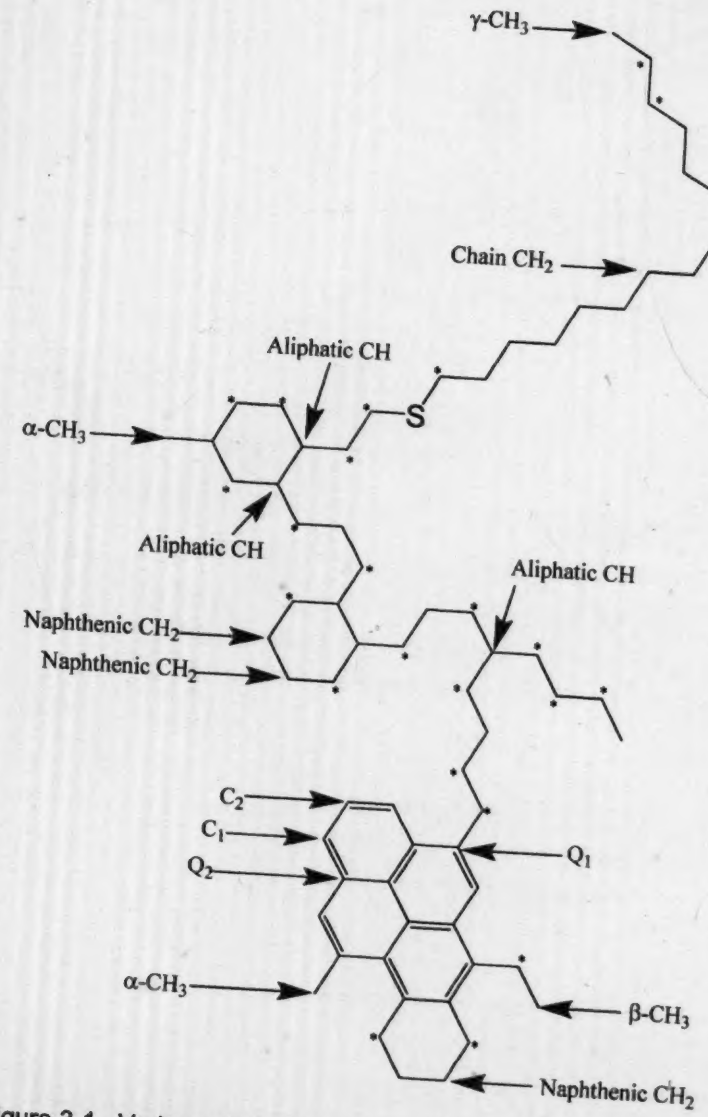


Figure 3-1. Various aliphatic and aromatic carbons present in the asphaltene molecular representation.

Q_1 = Alkyl-substituted aromatic quaternary carbon
Q_2 = Bridgehead aromatic quaternary carbon
C_1 = Aromatic CH beside a Q_2 carbon
C_2 = All aromatic CH that are not a C_1 carbon
* = Other aliphatic carbon

Table 3-3. Definitions of NMR spectroscopy determined chemical groups.

Carbon Chemical Species	Spectral Region
Alkyl-substituted aromatic quaternary carbon (Q_1)	HP1 + HP2 converted to C wt %
Bridgehead aromatic quaternary carbon (Q_2)	CA_2-Q_1
Aromatic α-CH_3	HP2 converted to C wt. %
Aliphatic CH	CP1 + CP2/2
Chain CH_2	CP4-CP8-CP5/2
Aromatic CH beside a Q_2 carbon (C_1)	HA1 converted to wt C%
All aromatic CH that are not a C_1 carbon (C_2)	HA2 converted to wt C%
Total α-CH_3	Ring α-CH_3 + Aromatic α-CH_3 + Aliphatic α-CH_3

Table 3-4. Experimental properties and their respective errors of the Athabasca residue sample.

Property	Experimental Properties $\mu_n \pm \varepsilon_n$
Mol Weight (g/mol)	4190 ± 630
% C (wt %)	81.4 ± 1.6
% H (wt %)	8.45 ± 0.17
% N (wt %)	1.17 ± 0.02
% S (wt %)	7.95 ± 0.16
% O (wt %)	1.03 ± 0.02
V (PPM)	877 ± 18
Total Aromaticity (% mol C)	50.1 ± 3.0
Q_1 (% mol C)	10.4 ± 3.6
Q_2 (% mol C)	17.6 ± 3.3
C_1 (% mol C)	10.8 ± 3.3
C_2 (% mol C)	11.3 ± 2.3
Total Aliphatic Content (% mol C)	49.9 ± 3.0
Other aliphatic (% mol C)	17.7 ± 4.6
Total α-CH_3 (% mol C)	6.6 ± 0.4
β-CH_3 (% mol C)	1.3 ± 0.2
Chain CH_2 (% mol C)	7.7 ± 0.3
Aliphatic CH (% mol C)	9.8 ± 2.1
Naphthenic CH_2 (% mol C)	4.5 ± 0.3
γ-CH_3 (% mol C)	2.3 ± 0.2

3.3.3. Molecular Generation

Figure 3-2 summarizes the procedure used to construct the asphaltene representations. A more detailed description of the construction procedure is presented in Appendix B. Estimates of the concentrations of heteroatom and metal containing groups, including indole, dibenzothiophene, dibenzofuran, and porphyrin, were calculated using the elemental concentrations. It was assumed that 60% of the sulfur was associated with the aromatic rings and the balance was associated with thioether bridges.[7] All the nitrogen was assigned to aromatic rings.[17] Analytical data regarding the distributions of oxygen and vanadium were not available. Therefore, it was assumed that all the oxygen was present in aromatic rings and all the vanadium was present within porphyrin rings. The aromatic ring groups used in the asphaltene representation are presented in Figure 3-3. Aromatic groups that contained five, six, and seven rings were also considered. However, a trial and error procedure revealed that the presence of these groups decreased the molecular representation fit to the experimental data. Thus, it was decided not to include these groups in the molecular representations. Likewise, it was determined that molecular representations created using dibenzothiophene and dibenzofuran had a better fit to the experimental data, compared to representations created that only used benzothiophene and benzofuran. Phenylnaphthalene and biphenylnaphthalene were included to allow the formation of larger aromatic groups.

The concentrations of benzene, naphthalene, phenanthrene, phenylnaphthalene, biphenylnaphthalene and pyrene were optimized using nonlinear optimization. The concentration of each of these aromatic groups was optimized so that the total Q_1, Q_2, C_1 and C_2 concentrations calculated from all of the aromatic groups were consistent with the analytical values obtained from NMR spectroscopy. Half of the naphthenic carbons were assumed to be associated with aromatic rings and half were associated with aliphatic carbons (Figure 3-1). It was assumed that one half of terminal α-CH_3 carbon was

associated with aromatics, one quarter associated with chain methyne carbons, and one quarter associated with naphthenic rings. The aliphatic groups used in the asphaltene representation are presented in Figure 3-1.

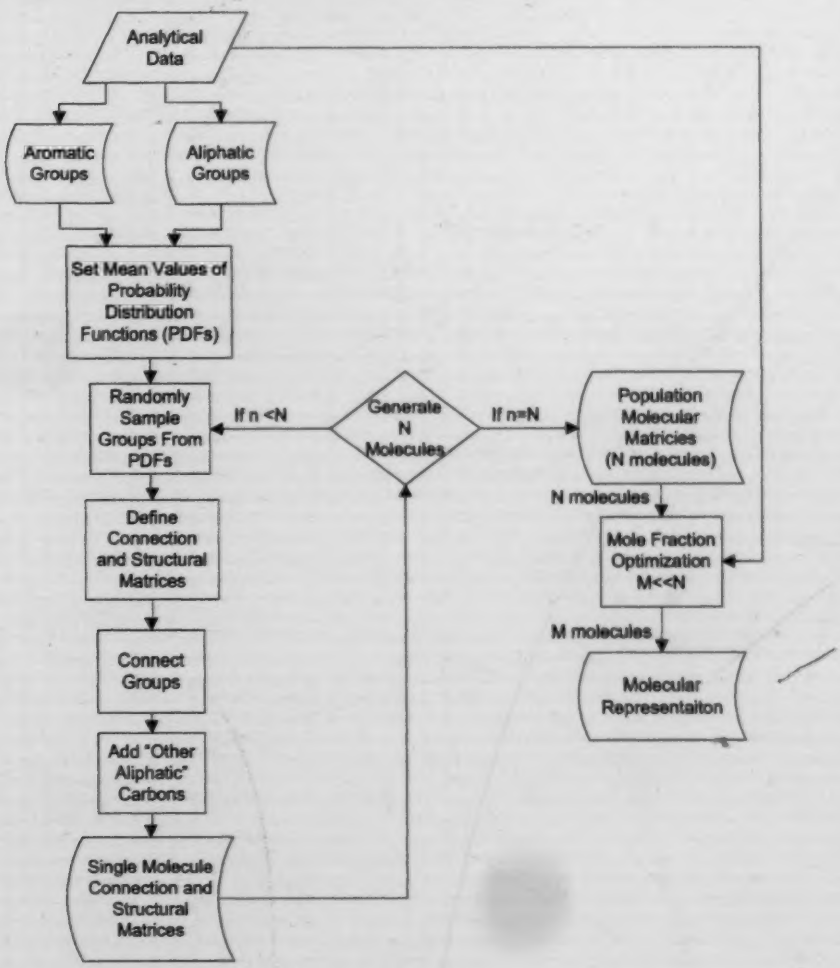


Figure 3-2. Flowchart for construction of a molecular representation using the Monte Carlo construction method.

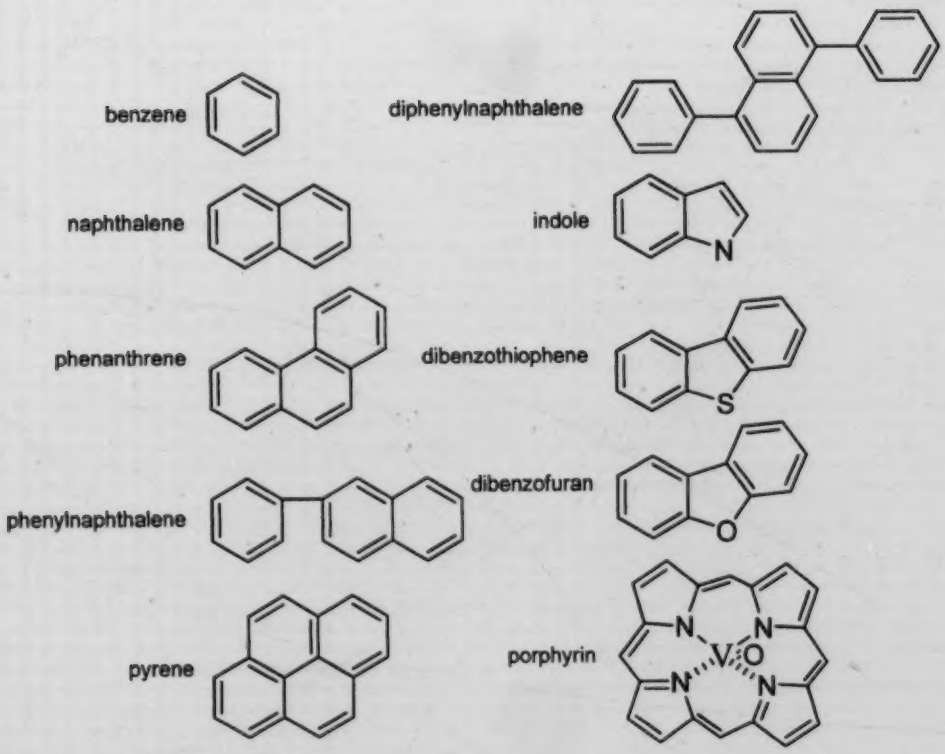


Figure 3-3. The aromatic groups present in the asphaltene representation.

Probability distribution functions (PDFs) were applied to each of the groups except for "other aliphatic" carbon. PDFs were also assigned to Q_1 carbon. The γ-distribution was chosen for the PDFs following Campbell and Klein.[4] The γ-distribution can be used to fit both skewed and normally distributed data. No data regarding the shape of the group distributions were available. Therefore, standard cumulative γ-distributions were used. However, if distribution data became available, it could be easily incorporated. A multi-step process was used to sample from the PDF from each group. In the first step, a random number was generated between 0 and 1. The number of groups corresponding to the random number was then calculated. The resulting sampled number was a positive real number with a whole number part and a decimal part. For construction of a given molecule, only whole numbers of groups are possible. A step function was then created with a mean set to the decimal part of the sampled number. A random number between 0 and 1 was then generated. If the random number was less than or equal to the decimal part, the output from the step function was 1. If the random number was greater than the decimal part, the output from the step function was 0. The output from the step function was then added to the whole number part to obtain the whole number of chemical groups of a given type.

Two matrices were used to define each molecule, one to specify the basic structures included in the molecule, and the other to specify how they were connected. Appendix C.1. discusses the matrix storage methodology in detail. A structural matrix was used to define all of the groups found in a molecule and the number of each type of group. A connection matrix was used to document all the internal connections within a molecule.

To create a molecule, the PDFs associated with the aromatic groups were randomly sampled. Several different aromatic groups would be simultaneously

generated for a given molecule. For each aromatic group that was generated, additional aromatic groups would be attached to each by sampling the PDFs associated with α-CH_3, β-CH_3, and Q_1 carbon. Naphthenic rings associated with aromatic groups were subsequently sampled and randomly assigned to specific aromatic groups. Each aromatic group was allowed to have $2n + 1$ groups attached to it, where n represents the number of rings in the aromatic group. If this constraint was not met, the attached groups were resampled. The chain CH_2, aliphatic CH, thioether, and aliphatic-associated naphthenic ring groups were then sampled.

Once all of the aromatic groups and substituents were sampled, the structural information was passed to a connection algorithm. To allow for aromatic groups larger than those presented in Figure 3, aromatic groups were allowed to connect to a maximum of one other aromatic group through a biphenyl bridge. After the aromatic and aliphatic groups were connected together, γ-CH_3 carbons were added at the end of aliphatic chains. Two "other aliphatic" carbons were added to the aromatic end of each aliphatic chain connected to an aromatic group, in order to represent the NMR spectroscopy data. The specific aromatic carbons involved in connections with paraffin chains, aromatic-associated naphthenic rings, aromatic α-CH_3, and β-CH_3 carbon were then determined and documented in the connection and structural matrices. At this point the structural and connection matrices were fully defined for a single molecule. Additional molecules were created by repeating the sampling and connection steps in the algorithm.

3.3.4. **Optimization**

The goal of this work was to represent an asphaltene fraction with a minimum number of molecules that were consistent with experimental data. With the procedure a relatively large number of molecules were generated, and then nonlinear optimization was used to select the best molecules from this

population. An L-1, absolute value, objective function was used for the optimizations (Equations 3-1 and 3-2):

$$F = \sum_{n=1}^{7} \left| \frac{\mu'_n - \mu_n}{\varepsilon_n} \right| + \frac{1}{11} \sum_{n=8}^{18} \left| \frac{\mu'_n - \mu_n}{\varepsilon_n} \right| + J \qquad (3\text{-}1)$$

where:

$$\mu'_n = \sum_{i=1}^{M} x_i \mu'_{n,i} \qquad (3\text{-}2)$$

x_i is the mole fraction of each molecule in a set, $\mu'_{n,i}$ is the value of property "n" from the molecule refers to the mean calculated value of property "n" from the molecular representation, ε_n is the experimentally determined value of property "n", ε_n is the calculated error of property "n" (Table 3-4), and "J" is a scaling factor that is required to keep aromaticity within experimental error boundaries. Equations 3-3 and 3-4 are equations used to calculate "J":

$$J = 1000000 \quad if \quad \left| \mu'_{aromatic} - (\mu_8 + \mu_9 + \mu_{10} + \mu_{11}) \right| \geq 1.5 \qquad (3\text{-}3)$$

$$J = 0 \quad if \quad \left| \mu'_{aromatic} - (\mu_8 + \mu_9 + \mu_{10} + \mu_{11}) \right| < 1.5 \qquad (3\text{-}4)$$

where $\mu'_{aromatic}$ is the aromaticity calculated from the molecular representation,

Both least squares and L-1 optimization methods were evaluated.[18] L-1 optimization was found to yield superior results when both methods used the same number of iterations for each optimization step. A sequential optimization scheme was used to select molecules from the Monte Carlo set. The objective function was evaluated for each of the molecules in the population. The molecule that gave the lowest value for the objective function was then selected. In the next step, all combinations between the selected molecule and all the other molecules in the population were examined. For each pair of molecules, nonlinear optimization was used to optimize the mole fractions for each molecule to match the experimental data (Equation 3-1). The pair with the lowest associated value of the objective function was then selected. The algorithm

continued adding molecules into the representation until the objective function reached a constant value with a tolerance of 1%.

In order to visualize the output of the model, the visualization algorithm used to transform the connection and structural matrices into "Mol" files (for a detailed description of the visualization algorithm see Appendix C.2.). This algorithm analyzed the connection matrix associated with a molecule and then assigned an unique (x,y) coordinate to each atom in the molecule. The connection information and the coordinates were then used to create "Mol" files for visualization or further refinement.

3.4. Results

3.4.1. Molecular Representation Properties

The data in Tables 3-5 and 3-6 present the optimized asphaltene representation consisting of molecules from a starting population of 100. The calculated aromaticity and elemental compositions of the molecular representation matched the experimental values to within the range of experimental error. Molecular weight had a virtually no deviation from the value determined using VPO. From previous experience using VPO on various petroleum residues, an error of at least 15% is normally associated with these measurements.

Table 3-6 lists the molecular weights and the elemental compositions of each of the six molecules. Molecules 1, 2, and 3 from Table 3-6 are shown in Figures 3-4, 3-5 and 3-6 respectively. All of the elements were modeled within experimental error (Table 3-5). Carbon content was relatively constant between the different molecules in the representation, varying between 79.5% and 82.2%. Hydrogen content was proportionally more variable: 8.2% to 8.8%. The

heteroatoms had even greater variability: 1.0-1.8% for nitrogen, 0.3-2.0% for oxygen, and 5.5-9.6% for sulfur. The majority of the groups were calculated to within experimental error (Table 3-5). γ-CH_3 was almost represented within the experimental error. "Other aliphatic", naphthenic CH_2, and chain CH_2 were not fit within the experimental error of the data or the replicate standard deviation.

Table 3-5. Comparison between experimental and calculated properties of the asphaltene molecular representation[a].

Property	Objective Function Index	Experimental Properties $\mu_n \pm \varepsilon_n$	Calculated Properties[b]	% Error	Replicate Standard Deviation[c]	Replicate Standard Deviation (%)
Mol Weight (g/mol)	1	4190 ± 630	4186	0.1	165.7	4.0
% C (wt %)	2	81.4 ± 1.6	81.3	0.1	0.1	0.1
% H (wt %)	3	8.45 ± 0.17	8.46	0.1	0.1	1.1
% N (wt %)	4	1.17 ± 0.02	1.17	0.0	0.0	0.4
% S (wt %)	5	7.95 ± 0.16	7.95	0.0	0.0	0.3
% O (wt %)	6	1.03 ± 0.02	1.03	0.0	0.0	0.0
V (PPM)	7	877 ± 18	877	0.0	0.0	0.0
Total Aromaticity (% mol C)		50.1 ± 3.0	48.3	3.5	1.9	3.7
Q_1 (% mol C)	8	10.4 ± 3.6	12.0	21.2	0.7	6.9
Q_2 (% mol C)	9	17.6 ± 3.3	16.8	6.3	0.5	2.7
C_1 (% mol C)	10	10.8 ± 3.3	8.5	28.7	0.7	6.6
C_2 (% mol C)	11	11.3 ± 2.3	11.0	1.8	1.2	10.8
Total Aliphatic Content (% mol C)		49.9 ± 3.0	51.7	3.52	1.9	3.7
Other aliphatic (% mol C)	12	17.7 ± 4.6	24.4	37.9	0.3	1.7
Total α-CH_3 (% mol C)	13	6.6 ± 0.4	6.5	1.1	0.4	6.2
β-CH_3 (% mol C)	14	1.3 ± 0.2	1.1	12.3	0.1	8.8
Chain CH_2 (% mol C)	15	7.7 ± 0.3	6.3	17.9	0.9	11.6
Aliphatic CH (% mol C)	16	9.8 ± 2.1	9.4	4.0	0.8	7.8
Naphthenic CH_2 (% mol C)	17	4.5 ± 0.3	2	55.6	0.6	13.8
γ-CH_3 (% mol C)	18	2.3 ± 0.2	1.8	21.7	0.2	9.1

a 100 molecules formed the initial population. Six molecules were present in the final molecular representation. *b* Calculated properties from a single representation. *c* Standard deviation of five replicate representations from independent construction and optimization calculations.

Table 3-6. Calculated elemental compositions of the six molecules in the asphaltene representation. 100 molecules formed the initial population. Six molecules were present in the final molecular representation.

			Elemental Composition (wt %)					
Molecule #	mol %	Mol Wt. (g/mol)	C	H	N	O	S	V
1	34.7	4133	82.2	8.2	1.0	7.0	1.5	0
2	19.2	3476	79.5	8.8	1.6	9.2	0.9	0
3	7.2	4705	81.2	8.5	1.8	5.5	2.0	1.1
4	20.2	4332	80.7	8.4	1.0	9.6	0.4	0
5	8.8	5120	81.6	8.8	1.1	8.1	0.3	0
6	9.9	4244	82.1	8.6	1.0	7.6	0.8	0

3.4.2. The Effect of Varying the Molecule Population Size

In order to examine the ability of molecular representations to fit the experimental data, independent populations of 25, 50, and 100, and 150 molecules were created. Each population was optimized using the sequential optimization approach. For each independent population, the objective function decreased as the number of molecules in the representation increased from one to five (Figure 3-7). Thus, a "high quality" representation required at least five molecules.

3.4.3. The Effect of Varying Molecular Weight

Because controversies exist concerning the experimental measurement of the average molecular weight of asphaltenes, the sensitivity of the molecular representations to molecular weight was examined. Two different target molecular weights were selected: one half of the VPO determined molecular weight (2055 g/mol) and one quarter of the VPO determined molecular weight (1028 g/mol). The data in Table 3-7 show the comparison between the calculated properties at different target molecular weights and the experimental data. The molecular weight did not affect the ability of this method to represent the data.

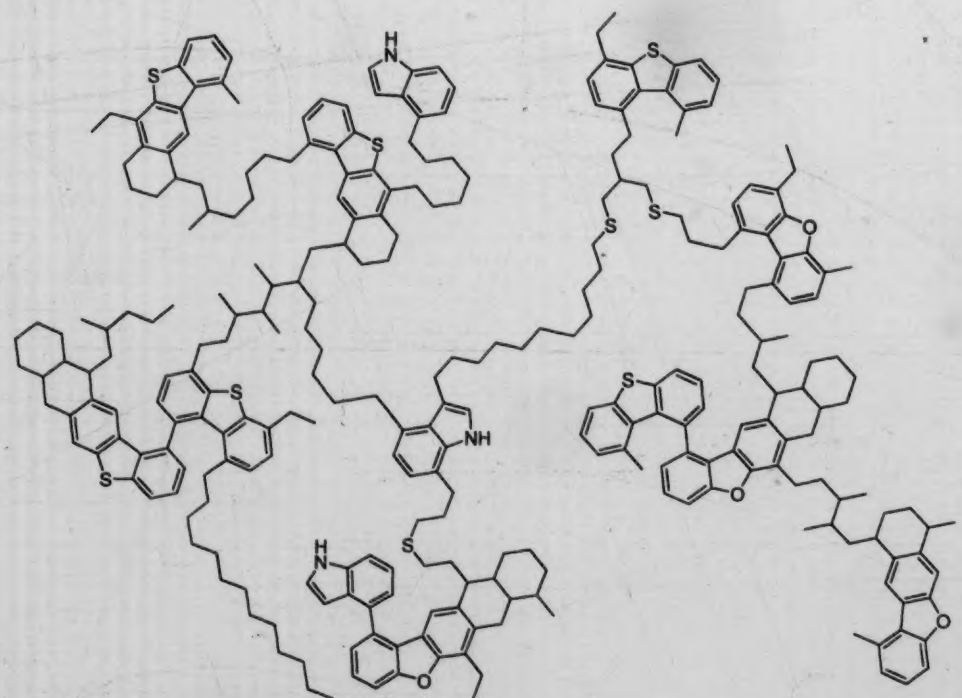


Figure 3-4. $C_{283}H_{337}N_3O_4S_9$ Mol. Wt.: 4133 g/mol.

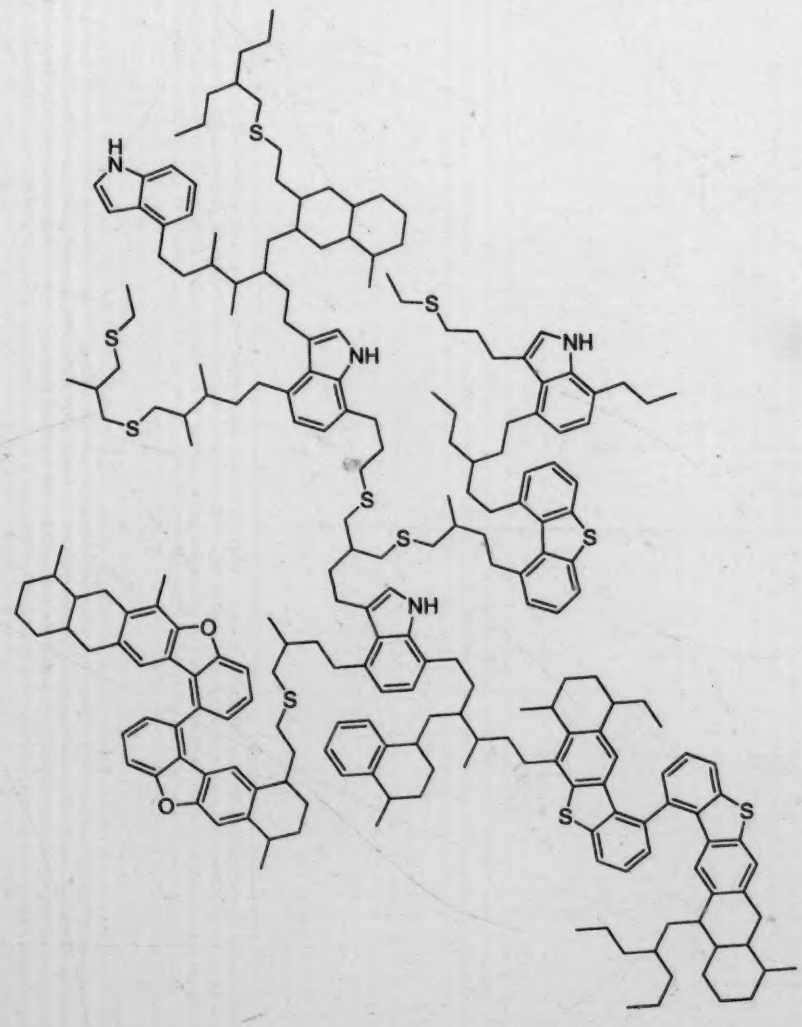


Figure 3-5. $C_{230}H_{302}N_4O_2S_{10}$.Wt.: 3476 g/mol.

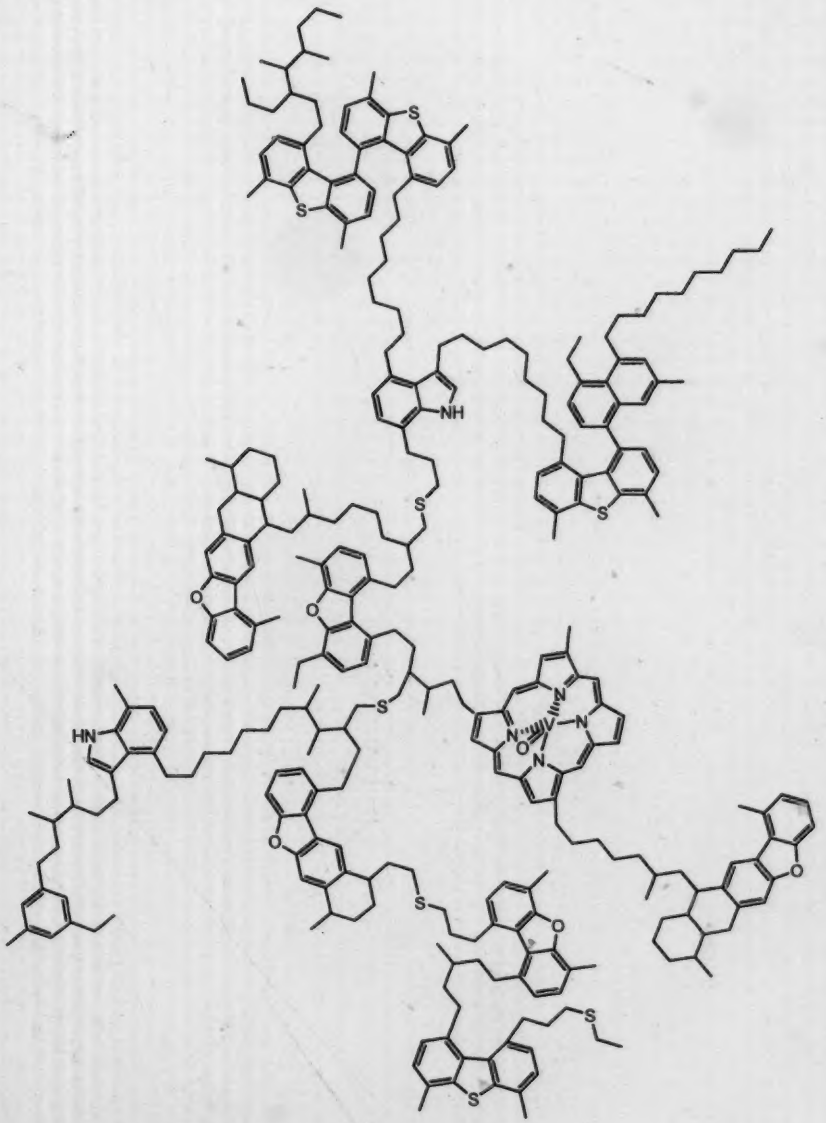


Figure 3-6. $C_{318}H_{395}N_6O_6S_8V$ Mol. Wt.: 4705 g/mol.

Table 3-7. Calculated properties of asphaltene representations of different target molecular weights. 100 molecules formed the initial population. Six molecules were present in the final molecular representation.

		Target Molecular Weight (g/mol)		
Property	Experimental	4190	2055	1028
Calculated MW (g/mol)	4190	4678	2140	1240
C (wt %)	81.4	81.4	81.4	81.4
H (wt %)	8.45	8.45	8.45	8.45
N (wt %)	1.17	1.17	1.17	1.17
S (wt %)	7.95	7.95	7.95	7.95
O (wt %)	1.03	1.03	1.03	1.03
Aromaticity (% mol C)	50.1	50.2	49.8	49.8

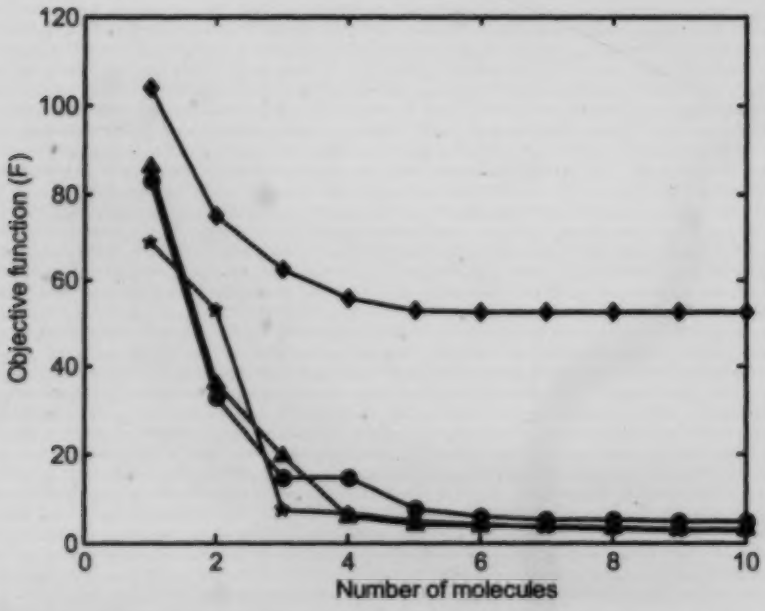


Figure 3-7. Sequential optimizations of populations of 25 (♦), 50 (▲), 100 (*), and 150 (•) asphaltene molecules.

3.5. Discussion

This study provides the first quantitative molecular representation of asphaltenes as archipelago structures. All of the structures generated in this study (Figures 3-4, 3-5, and 3-6) are composed of islands of relatively small aromatic clusters connected together by aliphatic and sulfur side chains. These are not rigid structures, and considerable folding would be expected in an appropriate solvent environment. The inclusion of phenylnaphthalene and diphenylnaphthalene groups, as shown in Figure 3-1, allowed aromatic clusters containing up to five biphenyl bridges to be generated by the construction algorithm. When the number of biphenyl connections allowed for each aromatic group was unconstrained, the quality of the molecular representations did not improve. This observation suggests that large aromatic clusters were not consistent with the NMR spectroscopy data.

The asphaltene molecular representation presented in this paper demonstrates the importance of utilizing both ^{1}H and ^{13}C-NMR spectroscopy to analyze asphaltene chemistry. If only ^{1}H-NMR spectroscopy had been used, there would have been less variety of aromatic and aliphatic primary structures. The specific concentrations of quaternary and protonated aromatic carbons would not have been differentiated, therefore, definition and modeling of the aromatic groups would have been more ambiguous. Also, if ^{13}C-NMR spectroscopy had not been used, the naphthenic content would have been unknown, while the concentrations of groups including aliphatic CH and β-CH_3 would not have been quantified. Thus, since both ^{13}C and ^{1}H-NMR spectroscopy data were used in concert, a more chemically detailed molecular representation was created.

With the NMR spectral analysis of petroleum fractions, it is often difficult to assign peaks to specific groups. The observed peaks can be relatively wide and there is considerable overlap of resonances. In this study there was a relatively undefined group of carbon species called "other aliphatic" carbons, shown in Figure 3-1, that posed a problem. Since "other aliphatic" carbons can show up in many different locations in a molecule, they did not have PDFs associated with them during the molecular construction process. They were added to each molecule after the more defined aromatic and aliphatic groups were sampled and connected. In the future, we plan to resolve the NMR spectra further so that all of the groups currently classified as "other aliphatics" are uniquely identified.

Sources of error in the representation originate from both the determination of the experimental property concentrations and from the inherent variability associated with Monte-Carlo sampling methods. As can be seen in Table 3-5, all of the elemental compositions were predicted to within the experimental error associated with elemental analysis. Given the errors associated with the NMR spectroscopic analysis of petroleum fractions, it is difficult to create asphaltene molecular representations that precisely match the quantities of all carbon types. The model predicted total aromaticity and associated aromatic subgroups (Q_1, Q_2, C_1 and C_2) to within the experimental error. Total aliphatic content was also predicted to within experimental error. There was more variability, however, in the error associated with some of the constituent aliphatic subgroups. Specifically, the most difficulty was encountered with matching the concentration of the "other aliphatic" group. Every time a representation is created using the Monte-Carlo construction technique the results change, and replicate calculations give different deviations from the experimental data. . For the majority of the properties, the standard deviation associated with replicates was less than 10%. More important, the variability associated with generation of the representation was within the experimental error of each of the measured properties.

Another source of error can come from bias in the model. To test for biases, a set of hypothetical structures is generated and the calculated properties are compared with the experimental data. Consistent biases between the model predictions and the experimental data are indications that the construction algorithm can be improved. For example, using this type of testing, we found that two-ring structures rings (e.g. decahydronaphthalenes) should be included in the asphaltene structures rather than only single-ring naphthenic groups. When only single naphthenic rings were included hydrogen content was too high, while carbon content was too low when the model was forced to fit the aromaticity data. The introduction of two-ring naphthenes solved this problem. A second example of an observed bias was a consistently high concentration of other aliphatic groups. This bias was addressed by changing how the aliphatic chains were connected to the hydroaromatic groups in the asphaltenes structures. The quality of the molecular representations improved when aliphatic groups were allowed to connect not only to the aromatic portions of the hydroaromatic groups but also to the aliphatic portions.

The algorithms and matrix representations used in this study are highly versatile and easily expandable. The structural and connection matrix framework can easily incorporate new chemical structures. In this study the relative ratio of basic to nonbasic nitrogen containing rings was not available. As more or improved analytical data comes available, it can readily be incorporated into the sampling, matrix, and construction methodology used in this study.

A petroleum residue fraction contains a vast number of individual molecules that will vary considerably in both the structural groups they contain and the relative abundance of these groups. Many previous studies have used an average molecule approach to represent petroleum fractions.[1,2,8,19,20] All of the sequential optimizations in this study indicate that more accurate representations were obtained when more than a single molecule was used to represent an asphaltene fraction. When a limited amount of analytical or structural data is taken into account, a consistent representation may be possible using a single molecule. With an extended data set, however, many chemical characteristics must be included into the overall representation and no single molecule is consistent with all the data. From the results shown in Figure 3-7, we conclude that the inclusion of three molecules greatly improved the quality of the molecular representation. The best fit of the representation to the analytical data was obtained with sets of five or six molecules. As well, the best fit required an adequate number of molecules in the starting population. In Figure 3-7, optimal fits were obtained if 50 or more molecules were used in the starting population.

The archipelago framework of the asphaltene structures generated in this study will allow reaction models to be developed that generate products that are consistent with experimental data from real reactions.[21] The connection algorithm used in this study did not allow cross-linking within a molecule as has been shown in schematic representations in the literature.[8] Molecules generated in this study were composed of a branched but not cross-linked network of aromatic and aliphatic groups, as illustrated in Figures 3-4, 3-5, and 3-6. Currently, there are no analytical data that describes the occurrence of crosslinking in asphaltene molecules. If present, intramolecular cross-linking would be important to include particularly if these molecules are used for reaction studies because crosslinking would influence the decomposition of asphaltenes into fragments[22].

3.6. Conclusions

Asphaltene molecular representations with multiple bridges between aromatic groups were constructed using a Monte Carlo method. After an optimal set of molecules was selected, the molecular representation fit the elemental data and ^{13}C-NMR spectroscopy aromaticity within experimental error. The majority of carbon groups calculated from ^{1}H and ^{13}C-NMR spectroscopy were represented accurately. The results from the sequential optimization indicate that at least five molecules are needed to create an analytically consistent molecular representation. For the sequential optimization scheme to produce a "high quality" molecular representation, a starting population of at least 50 molecules is required.

3.7. References

1. Brown, J.K. Ladner, W.R. Sheppard, N. *Fuel,* **1960**, *39*, 79-86.
2. Brown, J.K. Ladner, W.R. Sheppard, N. *Fuel*, **1960**, *39*, 87-96.
3. Neurock, M.; Nigam, A.; Trauth, D.; Klein, M.T. *Chem. Eng. Sci.* **1994**, *49*, 4153- 4177.
4. Campbell, D M.; Klein, M.T. *Applied Cat. A: General* **1997**, *160*, 41-54.
5. Joshi, P. Klein, M.T. *Rev. Process. Chem. and Eng.* **1998**, *1*, 111-140.
6. Strausz, O.P.; Peng, P. Murgich, J. *Energy Fuels* **2002**, *16*, 809-822.
7. Strausz, O.P.; Mojelsky, T.W.; Faraji, F.; Lown, E.M. *Energy Fuels* **1999**, *13(2)*, 207-227.
8. Strausz, O.P.; Mojelsky, T. W.; Lown, E. M. *Fuel,* **1992**, *71*, 1355-1363.
9. Strausz, O. P.; Mojelsky, T. W.; Lown, E. M.; Kowalewski, I.; Behar, F.; *Energy Fuels* **1999**, *13(2)*, 228-247.
10. Peng, P.; Morales-Izquierdo, A.; Hogg, A.; Strauz, O.P. *Energy Fuels* **1997**, *11,* 1171-1187.
11. Chung, K.H.; Xu, C.M.; Hu, X.Y.; Wang, R.N. *Oil Gas J.* **1997**, *95(1),* 66-70.
12. Thiel, J.; Gray, M.R. *AOSTRA J. Res.* **1988**, *4*, 63-73.
13. Snape, C.E.; Ladner, W.R.; Bartle, K.D. *Anal. Chem.* **1979**, *51(13),* 2189-2198.
14. Sarpal, A.S.; Kapur, G.S.; Chopra, A.; Jain, S.K.; Srivastava, S.P.; Bhatnagar, A. K. *Fuel* **1996**, *75*, 483-490.
15. Netzel, D.A. *Anal Chem.* **1987**, *59*, 1775-1779.
16. Kotlyar, L.S.; Morat, C.; Ripmeester, J.A. *Fuel* **1991**, *70(1)*, 90-94.
17. Jacobson, J.M.; Gray, M.R., *Fuel* **1987**, *66*, 749-752.
18. Watson, G.A.; Yiu, K.F.C. *BIT*, **1991**, *31(4)*, 697-710.
19 Groenzin, H.; Mullins, O.C. *Energy Fuels* **2000**, *14*, 677-684.
20 Rogel, E. Carbognani, L.; *Energy Fuels,* **2003**, *17,* 378-386.
21. Gray, M. R. *Energy & Fuels.* **2003**, *17(6)*, 1566-1569.
22. Solomon, P.R.; Hamblen, D.G.; Yu, Z.Z.; Serio, M.A.; *Fuel,* **1990**, *69*, 754-763.

Chapter 4. Quantitative Molecular Representations of Bitumen Residue Fractions

4.1. Introduction

In Chapter 3, a stochastic method was developed that created Athabasca asphaltene molecular representations. In this section, the methodology is applied to other Athabasca bitumen residue fractions. An important question arises: do molecular representations improve our understanding of bitumen structural chemistry beyond what is achieved from raw data alone?

This chapter has two objectives:

1. Present molecular representations created for other Athabasca bitumen residue fractions.
2. Analyze molecules representing the molecular representations to identify trends and structural information that are not implicit in the raw data used to construct the molecular representations.

4.1.1 The Relationship Between Structure and Function

The structure of a molecule ultimately determines its function and physical properties. Thus, by analyzing the structure of a molecule, we can gain insights into the function of a molecule. However, it is often difficult to describe a structure in quantitative terms. When one looks at two molecules, it is often apparent that the two molecules are different, but it is difficult to explain why or how they differ. For petroleum molecules, this is indeed the situation. All petroleum molecules contain aliphatic and/or aromatic groups in various combinations and arrangements. At first glance, many petroleum molecules look virtually identical: islands of aromatics, naphthenic rings, and aliphatic chains.

However, a closer look will make it apparent that the molecules differ in many ways such as in the length of sidechains, number of side chains, size of aromatics, number of aromatic groups, etc. Also, one must be very careful when choosing a basis when comparing molecules of different sizes or molecular weights. For example, one might be analyzing two residue molecules; one molecule with a molecular weight of 600 g/mol, and another molecule with a molecular weight of 1200 g/mol that is a dimer composed of two clones of the first molecule. If the molecules are described as containing a given number of predefined groups, the second dimmer molecule will contain exactly twice as many groups as the first molecule in absolute terms. Thus, on an absolute basis, the two molecules are very different – the higher molecular weight molecule contains twice as many groups as the first molecule. However, if the various groups are expressed per carbon, or per g of molecule, or per mol of molecule, the various group concentrations will be identical. Thus, on a constant mass or mol basis it is apparent that the two molecules are very similar. It is apparent that one must use either a consistent mass or consistent mol basis when comparing molecules.

Petroleum scientists and engineers are most interested in the bulk physical properties of petroleums: viscosity, solubility, density, heats of reaction, etc. However, in most cases, the scientific fundamentals at the molecular level that influence the bulk physical properties of petroleum are poorly understood. At the molecular level, it is reasonable that certain structural indexes will significantly influence bulk properties. Viscosity is a measure of the resistance to flow of a liquid. Thus, structural elements that provide flow resistance – such as long aromatic side chains - would be expected to profoundly influence viscosity. The first step in studying the relationship between molecular properties and bulk physical properties is to define the molecular properties of interest. Since petroleum is composed of aliphatic and aromatic structural groups, it is most reasonable to use indexes that characterize the distribution, size and shape of the aromatic and aliphatic groups. In the case of the aromatics, indexes that

measure number of aromatic rings/cluster, and number of aromatic rings, seem likely to provide the most insight into petroleum structure. Likewise, in the case of the aliphatics, indexes that will give insights into petroleum structure and can potentially give insights into the relationship between structure and petroleum properties include: length of branches and sidechains (both originating from aromatic and aliphatic groups), length of aliphatic chains between aromatics, number of naphthenic rings, and number of naphthenic rings per cluster.

Obtaining detailed structural information of petroleum fractions is challenging. The collection of primary information such as elemental compositions is fairly straight forward. NMR spectroscopic analysis can yield further information regarding the types of aromatic and aliphatic structures and their environments. Translating all of the available analytical data into a form that can be used for property prediction and reactivity analysis is challenging. The main challenge is converting the available analytical data into a molecular form. Currently, there are no experimental techniques that yield information about the overall primary or secondary structures of petroleum molecules. Petroleum is a mixture of tens of thousands of unique molecules. In the cases of the highest molecular weight petroleum fractions including residues and asphaltenes, individual molecules cannot be isolated and characterized. When a homogeneous source of molecules can be produced, such as with DNA and proteins, detailed primary structure information can be obtained using analytical techniques such as X-ray diffraction. With a mixture of molecules, we must utilize techniques such as NMR spectroscopy or mass spectroscopy to yield information about the chemical structure of a petroleum fraction. This information is for the entire fraction, not for individual molecules. In the past, both graphical[1] and Monte-Carlo techniques[2,3] have been used to create representatives of the molecules in a petroleum fraction. Detailed chemical representations can be generated by combining data from a wide variety of analytical techniques with a Monte-Carlo algorithm to construct molecules and then using mathematical

optimization to select a final representation. This approach has been used to predict the structure of a variety of petroleum fractions.[3]

4.1.2. Solubility Parameter Modelling of Bitumens

Hildebrand and Scott first described the solubility parameter.[4] Solubility paramaters describe the interactions between molecules in condensed materials. They can be defined as the difference in energy between the internal energy of the condensed material and that of an ideal gas of the same material at the same temperature. The solubility parameter defined by Hildebrand is:

$$\delta = \sqrt{c} = \sqrt{-\frac{U}{V}} \quad (4\text{-}1)$$

Where c is the cohesive energy density, U is the molar internal energy, V is the molar volume.

The Hildebrand solubility parameter is intended for non-polar, non-associating systems.[4] Multicomponent solubility parameters have been proposed to deal with systems characterized by high degrees of hydrogen bonding and polar interactions. The Hansen solubility parameter is a three component example that was formulated to account for dispersive forces, polar interactions, and hydrogen bonding[4]:

$$\delta_t^2 = \delta_d^2 + \delta_p^2 + \delta_h^2 \quad (4\text{-}2)$$

Where δ_t is the overall Hildebrand solubility parameter, δ_d is the dispersive component, δ_p is the polar component, and $\delta_h$is the hydrogen bonding component.

Bitumens are mixtures of molecules. Fractionation techniques such as SARA and supercritical extraction separate bitumens on the basis of solubility.

The solubility of bitumens in different solvents can be successfully predicted by both the Hildebrand and Hansen solubility parameter models.[4,5] Bitumen solubility parameters can easily be calculated from experimental data.

4.2. Materials and Methods

4.2.1. Residue Molecular Representations

Ten quantitative molecular representations were created for fractionated material supplied by Syncrude Canada Ltd., which was obtained from the supercritical fluid extraction of an Athabasca vacuum residue (+524℃)[6,7,8]. The fractions have been previously characterized and include these analyses: elemental composition, molecular weight, aromaticity and SARA fractionation. Selected properties of the fractions are presented in Table 4-1.

Table 4-1. Properties of fractionated Athabasca vacuum residue.

Fraction	1	2	3	4	5	6	7	8	9	10	Error (%)
Mass %*	12.7	9.8	7.6	10.6	6.5	4.4	3.3	2.6	1.1	41.4	-
MW	506	755	711	800	825	948	1138	1210	1520	4190/ 1800‡	10.0
% Aromaticity	27.5	29.7	28	34	36	40	40	42	46	50	3.0
% Aromatic S**	64.7	67.0	69.2	71.4	73.7	73.5	73.3	73.1	71.7	70.4	5.0
% Saturate	26.9	16.4	9.68	-	-	-	-	-	-	-	6.0
H/C	1.61	1.59	1.56	1.52	1.47	1.43	1.39	1.39	1.37	1.23	-

*[7], ** [8], ‡ [9]

Fractions 1 to 9 were heavy maltenes. Saturates, mixed with aromatics, were found in Fractions 1 to 3. Fraction 10, the asphaltenes rich fraction, contained 88% by weight n-C_5 asphaltenes and the balance heavy maltenes. Quantitative ^{1}H and ^{13}C-NMR spectroscopy were used to further characterize the ten fractions. The details of the NMR spectroscopy data acquisition are described in Chapter three of this thesis. The NMR spectroscopic analyses and interpretations of these fractions were performed by Dr. Heather Dettman at NCUT. Previously published XPS data was used to model the relative distribution of aromatic and aliphatic sulfur[8]. The average molecular weights of the fractions were determined by VPO at 130 ℃ usin g o-dichlorobenzene as the solvent in order to minimize molecular associations. The reported molecular weight of 4190 g/mol for fraction 10 appeared to be high, so a value of 1800 g/mol that had been reported for Athabasca asphaltenes was used for construction of the representations.[9]

4.2.2. Objective Function Formulation

The Monte-Carlo procedure described in Chapter three of this thesis was used to construct the molecular representations of the ten fractions. The construction algorithm was modified to include both benzothiophene and dibenzothiphene aromatic groups. This modified algorithm was used to construct all ten molecular representations. The objective function that was used to optimise the molecular representations was slightly different – it contained variables associated with the concentrations of aliphatic sulfur and saturates. The objective function for fractions one to three, also contained a scaling factor for the aromatic/saturate distribution. The general L-1, absolute value objective function that was used to optimize the molecular representations for the ten fractions is:

$$F = \sum_{n=1}^{8} \left| \frac{\mu'_n - \mu_n}{\varepsilon_n} \right| + \frac{1}{11} \sum_{n=9}^{19} \left| \frac{\mu'_n - \mu_n}{\varepsilon_n} \right| + J + \mathrm{A} \qquad (4\text{-}3)$$

where:

$$\mu'_n = \sum_{i=1}^{M} x_i \mu'_{n,i} \qquad (4\text{-}4)$$

The optimized properties are listed in Table 4-2. x_i is the mole fraction of each molecule in a molecular representation, $\mu'_{n,i}$ is the value of property "n" from the molecular representation, μ_n is the experimentally determined value of property "n", ε_n is the estimated error of property "n", and "*J*" is a scaling factor that maintains aromaticity within experimental error. Likewise, "*A*" is a scaling factor to keep the aromatic/saturate distribution within experimental error. Equations 4-5 and 4-6 are equations used to calculate "*J*" while Equations 4-7 and 4-8 were used to calculate "*A*".

$$J = 1000000 \quad if \quad \left|\mu'_{aromatic} - Aro_{exp}\right| \geq 3.0 \qquad (4\text{-}5)$$

$$J = 0 \quad if \quad \left|\mu'_{aromatic} - Aro_{exp}\right| < 3.0 \qquad (4\text{-}6)$$

Where "*J*" is a scaling factor that maintains aromaticity within experimental error, $\mu'_{aromatic}$ is the aromaticity calculated from the molecular representation, and Aro_{exp} represents the experimentally determined aromaticity.

$$A = 10000 \quad \text{if} \quad \left|\text{Sat}_{calc} - \text{Sat}_{exp}\right| \geq 6 \qquad (4\text{-}7)$$

$$A = 0 \quad \text{if} \quad \left|\text{Sat}_{calc} - \text{Sat}_{exp}\right| < 6 \qquad (4\text{-}8)$$

Where "*A*" is a scaling factor to keep the aromatic/saturate distribution within experimental error, Sat_{calc} is the calculated saturate content of a molecular representation and Sat_{exp} is the experimentally determined saturate content.

Table 4-2. Optimized properties of the residue molecular representations.

Property #	Property
1	MW (g/mol)
2	% C
3	% H
4	% N
5	% S
6	% aliphatic S
7	% O
8	V (PPM)
9	%C Q_1
10	%C Q_2
11	%C C_A
12	%C C_B
13	%C Other Aliphatic
14	%C α-CH_3
15	%C β-CH_3
16	%C Chain methylene
17	%C Chain methyne
18	%C Naphthenic
19	%C γ-CH_3

4.3. Results and Discussion

Quantitative molecular representations were generated for each of the ten fractions of the vacuum residue. To create each fraction, initially 100 molecules were generated and then a limited number were selected though sequential optimization as described in Chapter three of this thesis (using Equations 4-3 to 4-8). Each molecule in the molecular representation was required to have a mole fraction of at least 0.01. In each case, six molecules were required to represent each fraction (Table 4-3).

Table 4-3. Detailed composition of the ten Athabasca residue molecular representations.

Fraction #	Molecule #	Concentration (mole % total molecules)	Molecular Weight (g/mol)	Total C	Total H	Total N	Total S	Total O	Total V
1	1	30.4	475	34	66	0	0	0	0
	2	20.0	597	39	64	0	2	0	0
	3	19.8	481	35	60	0	0	0	0
	4	12.5	555	37	46	0	2	0	0
	5	12.3	770	57	71	1	0	0	0
	6	4.9	749	53	80	0	1	0	0
2	1	37.1	687	48	78	0	1	0	0
	2	23.4	1083	75	103	1	2	0	0
	3	17.2	719	50	86	0	1	0	0
	4	11.2	513	36	48	0	1	0	0
	5	11.1	986	71	132	0	0	0	0
	6	0.0	1294	94	164	0	0	0	0
3	1	64.8	629	44	68	0	1	0	0
	2	13.7	720	48	65	1	2	0	0
	3	8.4	986	71	132	0	0	0	0
	4	7.7	692	47	81	1	1	0	0
	5	4.7	1076	75	110	0	2	0	0
	6	0.7	561	38	72	0	1	0	0
4	1	29.7	878	62	87	1	1	0	0
	2	25.0	926	64	92	0	2	0	0
	3	16.9	701	49	80	0	1	0	0
	4	13.1	641	45	68	0	1	0	0
	5	9.5	892	59	86	0	3	0	0
	6	5.8	703	52	78	0	0	0	0
5	1	40.0	800	56	81	1	1	0	0
	2	22.4	799	54	86	0	2	0	0
	3	14.2	1172	82	122	0	2	0	0
	4	13.0	952	66	94	0	2	0	0
	5	10.4	1162	80	104	0	3	0	0
	6	0.0	1022	71	104	0	2	0	0
6	1	44.2	762	51	71	1	2	0	0
	2	38.0	980	68	98	0	2	0	0
	3	11.2	1062	79	112	0	0	0	0
	4	4.8	1298	94	136	0	1	0	0
	5	1.6	1218	82	104	0	4	0	0
	6	0.3	1285	90	125	1	2	0	0
7	1	63.8	1347	92	131	1	3	0	0
	2	19.1	1214	84	108	0	3	0	0
	3	11.3	1152	86	118	0	0	0	0
	4	4.0	1179	82	115	1	2	0	0
	5	1.6	1298	90	120	0	3	0	0
	6	0.3	1477	102	141	1	3	0	0
8	1	38.1	1242	86	112	0	3	0	0
	2	33.1	1015	69	107	1	2	0	0
	3	20.9	1500	106	147	1	2	0	0
	4	5.7	1255	83	115	1	4	0	0
	5	1.8	1052	74	102	2	1	0	0
	6	0.4	1358	99	140	2	0	0	0
9	1	26.1	1621	117	150	0	2	0	0
	2	21.1	1173	79	113	1	3	0	0
	3	19.8	1151	78	103	1	3	0	0
	4	17.6	1067	71	103	1	3	0	0
	5	13.3	1475	103	127	1	3	0	0
	6	2.2	1281	91	109	1	2	0	0
10	1	50.4	1177	79	101	1	0	3	0
	2	23.0	2727	183	225	1	8	2	0
	3	15.6	2132	149	171	3	4	0	0
	4	7.8	1967	133	164	2	5	1	0
	5	3.0	2100	142	206	4	3	1	1
	6	0.0	2350	156	209	5	3	3	1

4.3.1. Molecular Representation Optimizations

As can be seen in Figure 4-1, the value of the objective function used to select the optimum molecules decreased with each additional molecule until six molecules were included in the population and the objective function reached a minimum. Figure 4-1 is for the Fraction 10 molecular representation. The same graphical trends were observed for the molecular representations of the other nine fractions. These results highlight the importance of including more than one molecule in a molecular representation. Many previous studies have represented petroleum fractions by using a single molecule.[10,11] Conversely, other previous studies have represented petroleum fractions using tens of thousands of molecules.[4] This work continues to demonstrate that it is critical to use more than a single molecule, but not necessary to use a very large population of molecules to represent a residue petroleum fraction. As more diverse types of data are included in a molecular representation for a petroleum fraction, more molecules will likely be required.

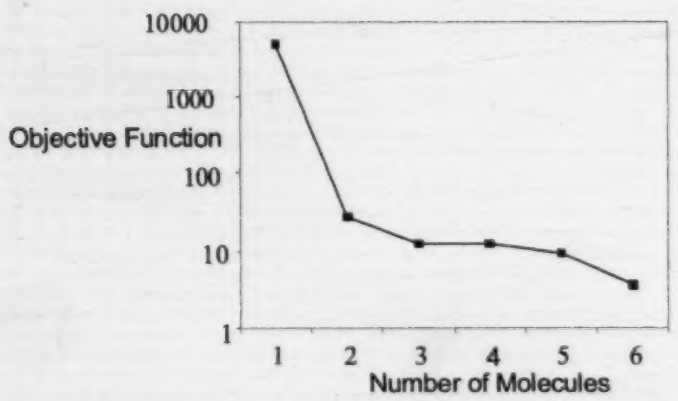


Figure 4-1. Sequential optimization for the creation of a molecular representation of Fraction 10.

4.3.2. Molecular Representation Performance

The calculated properties of each of the ten molecular representations are presented in Tables 4-4 to 4-7. Every fraction had the elemental concentrations, molecular weight, and total aromaticity modelled within experimental error. The molecular representations for fractions one, two and three had total deviations higher than those for fractions four through ten. However, the additional scaling factor for the saturate compositions complicated the objective functions of these representations. The aromatic sulfur contents of each of the ten molecular representations closely matched the experimental XPS data. For all of the ten fractions, "other aliphatic" carbon was significantly higher than the experimentally determined values. The molecular construction algorithm inserted "other aliphatic" carbons into the molecules, after all of the other groups were connected together. Thus, the construction algorithm did not have full control over these specific type of carbons. Additionally, "other aliphatic" carbons refer to many different types of carbon-as discussed in Chapter three. As a result, it is very difficult to improve the level of "other carbon" when it is difficult to locate the specific types of "other carbon" that are responsible for the deviation. As a result of the relatively high error in the "other carbon" content, the other carbon types have relatively higher errors - since the summation of all the carbon type percentages must be 100%. Except for "other aliphatic" carbon, the aliphatic carbon types had good agreement between the calculated and experimentally determined values. In the case of the aromatics, total aromaticity consistently had excellent agreement between the experimentally determined and calculated values. However, the agreement between the experimentally determined and calculated values for the individual aromatic groups - Q_1, Q_2, C_1, and C_2 - was variable from one fraction to another.

Table 4-4. Calculated properties of the molecular representations of Fractions 1-3.

Fraction #	1	1	1	2	2	2	3	3	3
	C*	E*	D* (%)	C	E	D (%)	C	E	D (%)
MW (g/mol)	561	506	10.8	800	755	6.0	697	711	2.0
% C	84.3	84.5	0.2	83.9	83.5	0.5	83.5	83.5	0.0
% H	11.4	11.5	0.5	11.2	11.2	0.0	11	11.0	0.0
% N	0.31	0.3	0.6	0.41	0.4	0.0	0.43	0.4	0.7
% S Total	4	4	0.0	4.5	4.5	0.0	5.1	5.0	2.0
% Aromatic S	64.4	64.7	0.3	67.0	67.0	0.0	71.3	69.2	2.1
%C Q_1	7.76	10	22.4	9.43	10.5	10.2	8.7	11.8	26.3
%C Q_2	7.2	6	20.0	7.83	9.2	14.9	7.94	4.9	62.0
%C C_A	3.9	4.5	13.3	1.9	4.2	54.8	3.34	4.9	31.8
%C C_B	5.8	7	17.1	7.93	5.8	36.7	5.33	6.4	16.7
% C Aromaticity	24.7	27.5	2.8	27.1	29.7	2.6	25.3	28.0	2.7
% C Saturate	30.4	26.9	3.5	11.1	16.4	5.3	9.1	9.7	0.5
%C Other Aliphatic	30.8	27	14.1	30.14	24.0	25.6	28.45	25.9	9.8
%C α-CH_3	7.2	9.9	27.3	9.62	10.4	7.5	9.3	9.0	3.3
%C Aro β-CH_3	1.3	1.7	23.5	0.82	1.5	45.3	0.5	1.9	73.7
%C Chain methylene	9.3	9.3	0.0	9	9.3	3.2	7.3	8.8	17.0
%C Chain methyne	15.2	13.9	9.4	13.7	14.0	2.1	17.4	16.0	8.7
%C Naphthenic	4.7	6.6	28.8	4.28	7.0	38.9	5.8	6.6	12.1
%C γ-CH_3	7.0	4.1	70.0	5.4	4.2	28.6	5.9	3.7	59.5

C=Calculated value from molecular representation; E=Experimentally determined value; D=Difference between the calculated and experimental values.

Table 4-5. Calculated properties of the molecular representations of Fractions 4-6.

Fraction #	4	4	4	5	5	5	6	6	6
	C	E	D (%)	C	E	D (%)	C	E	D (%)
MW (g/mol)	821	800	2.7	910	825	10.3	913	948	3.7
% C	83.7	84.0	0.3	83.2	83.0	0.2	83.2	84.0	1.0
% H	10.4	10.6	1.9	10.2	10.3	1.0	9.9	10.1	2.0
% N	0.5	0.5	0.6	0.616	0.6	0.0	0.68	0.7	0.1
% S Total	5.4	5.4	0.0	6.0	6.0	0.0	6.2	6.2	0.0
% Aromatic S	71.4	71.4	0.0	73.7	73.7	0.0	71.1	73.5	2.4
%C Q_1	10	11.8	15.3	11.0	12.5	12.2	11.1	12.6	11.9
%C Q_2	9.5	9.8	3.1	9.64	10.3	6.4	11.2	14.9	24.8
%C C_A	5.7	4.9	16.7	7.82	6.2	26.1	6.46	6.2	4.2
%C C_B	5.7	7.3	21.9	6.3	7.0	10.0	8.3	6.4	29.7
Aromaticity (% C)	30.9	33.8	2.9	34.7	36.0	1.3	37.1	40.1	3.0
%C Other Aliphatic	28.1	22.2	26.6	27.97	21.8	28.3	27.3	20.3	34.5
%C α-CH_3	5.1	8.2	37.8	6.99	8.2	14.8	6	8.8	31.8
%C Aro β-CH_3	1.2	1.2	0.8	1.3	1.6	18.8	0.9	1.6	43.8
%C Chain methylene	8.9	8.8	1.4	7.98	8.4	5.0	7.3	7.6	3.9
%C Chain methyne	14.4	16.5	12.7	12.53	15.1	17.0	12	11.7	2.6
%C Naphthenic	6.4	5.8	10.6	3.57	5.8	38.4	3.74	6.2	39.7
%C γ-CH_3	4.9	3.3	47.7	4.95	3.3	50.0	5.6	3.7	51.4

C=Calculated value from molecular representation; E=Experimentally determined value; D=Difference between the calculated and experimental values.

Table 4-6. Calculated properties of the molecular representations of Fractions 7-9.

Fraction #	7	7	7	8	8	8	9	9	9
	C	E	D (%)	C	E	D (%)	C	E	D (%)
MW (g/mol)	1292	1134	13.9	1218	1209.0	0.7	1310	1517	13.6
% C	83.1	83.0	0.1	83	83.0	0.0	83.2	82.5	0.8
% H	9.7	9.8	1.0	9.7	9.7	0.0	9.4	9.5	1.1
% N	0.737	0.7	0.4	0.737	0.7	0.0	0.79	0.8	0.0
% S Total	6.5	6.5	0.0	6.5	6.8	4.4	6.65	6.8	2.2
% Aromatic S	75.6	73.3	2.3	73.5	73.1	0.4	73.6	71.7	1.9
%C Q_1	11.6	12.8	9.4	11.37	12.4	8.3	12.99	11.6	12.0
%C Q_2	11.7	14.4	18.8	14.38	15.8	9.0	13.81	19.0	27.3
%C C_A	9.3	6.8	36.8	7	7.2	2.8	6.38	8.1	21.2
%C C_B	4.9	6.5	24.6	6.4	6.7	4.5	9.6	7.1	35.2
Aromaticity (% C)	37.5	40.5	3.0	39.2	42.1	3.0	42.8	45.8	3.0
%C Other Aliphatic	28.1	20.0	40.5	29.9	19.8	51.0	26.26	16.8	56.3
%C α-CH_3	5.5	8.1	32.1	6.1	7.9	22.8	5.3	7.4	28.4
%C Aro β-CH_3	1.6	1.9	15.8	1.6	1.9	15.8	1.9	1.7	11.8
%C Chain methylene	7.5	8.1	7.4	8	7.9	1.3	7.1	8.4	15.5
%C Chain methyne	11.4	12.5	8.8	7.4	10.5	29.5	8.83	11.4	22.5
%C Naphthenic	4.9	5.9	16.9	4.3	6.4	32.8	4.4	5.4	18.5
%C γ-CH_3	3.4	3.1	9.7	3.5	3.4	2.9	3.4	3.1	9.7

C=Calculated value from molecular representation; E=Experimentally determined value; D=Difference between the calculated and experimental values.

Table 4-7. Calculated properties of the Fraction 10 molecular representation.

Fraction #	10 C	10 E	10 D (%)
MW (g/mol)	1774	1800	1.5
% C	81.3	81.4	0.1
% H	8.5	8.5	0.0
% N	1.2	1.2	0.0
% S Total	8.0	8.0	0.1
% Aromatic S	72.4	70.4	2.0
%C Q_1	1.0	1.0	1.0
%C Q_2	0.1	0.1	0.1
%C C_A	12.8	10.4	23.2
%C C_B	16.7	17.6	5.3
Aromaticity (% C)	8.7	10.8	19.8
%C Other Aliphatic	9.5	11.3	16.2
%C α-CH_3	47.6	50.1	2.5
%C Aro β-CH_3	23.4	17.7	32.5
%C Chain methylene	5.6	6.6	15.8
%C Chain methyne	1.5	1.3	15.0
%C Naphthenic	7.1	7.7	8.4
%C γ-CH_3	9.6	9.8	1.8

C=Calculated value from molecular representation; E=Experimentally determined value; D=Difference between the calculated and experimental values.

4.3.3. Large Scale Optimization Results

Interesting results were observed when molecular representations that contained many molecules were created for Fraction 8 (Figure 4-2). In this particular case study, as with the other studies, 100 molecules were initially created. However, sequential optimization was used to create molecular representations that contained more than six molecules. The objective function between six and 23 molecules was relatively constant. However, once the molecular representations contained 24 or more molecules, the objective function increased significantly. Thus, in the entire population of molecules, there were a limited number of "good" molecules that allowed low objective function values to be obtained. This would be expected for a random construction methodology. A random construction methodology would be expected to construct a wide range of molecules: molecules that closely matched the analytical data-that had corresponding low objective function values, and molecules that deviated

significantly from the analytical data-that corresponded to high objective function values. Also, the optimization algorithm might not be capable of efficiently optimizing systems that contain a large number of variables. The objective function increase between 24 and 25 molecules followed by the decrease between 25 and 26 molecules, might be indicative of problems with the optimization algorithm.

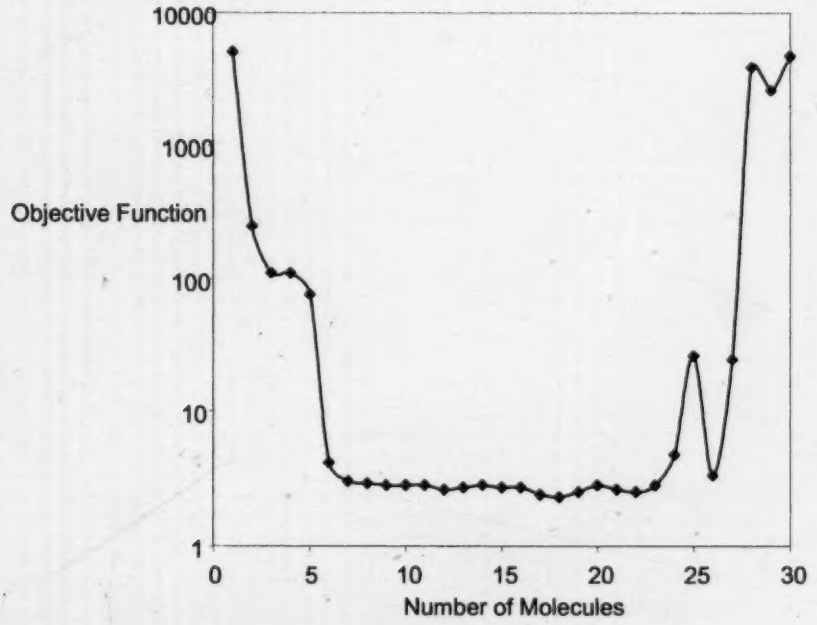


Figure 4-2. Large scale sequential optimization of Fraction 8.

4.3.4. Example Molecules

In the case of the high asphaltenes fraction - fraction 10 - six molecules were included in the representation with an average molecular weight of 1773 g/mol, ranging from 1177 to 2727 g/mol. In this study, aromatic sulphur was also included in the optimization. This allowed the ratio of aromatic to aliphatic sulfur to be modelled. An example of one of the molecules in Fraction 10 is

presented in Figure 4-2. The molecules all follow the archipelago model of construction with relatively small clusters of aromatics and naphthenic groups connected together by aliphatic carbon bridges. Thermally labile thioether groups were included in the bridging aliphatics.

The data for the molecular representations were generated from the analysis of fractions obtained by supercritical extraction of Athabasca vacuum residue. Usually this data is plotted against molecular weight or by cumulative mass yield of the fractions. Since the fractions were obtained by supercritical pentane extraction, an alternative method to analyse the data is by using solubility parameters. The Giddings equation (Equation 4-9) was used to calculate the solubility parameters of the solvent used to extract fractions one through nine. With solubility, "like dissolves like". Thus the solvent solubility parameter truly reflects the solubility characteristics of the bitumen fractions. The temperature and pressures that were used to produce the supercritical fractions were used to calculate the solvent gas density - (n-Pentane) –using the Peng Robinson equation of state for these nine fractions. Figure 4-3 presents the relationship between the solubility parameter and the molecular weight as calculated from VPO for each fraction except for the last fraction as previously discussed.

$$\delta=1.25P_C^{1/2}\rho_r \quad (4\text{-}9)$$

Where P_c is the solvent critical pressure and ρ_r is the reduced solvent density.

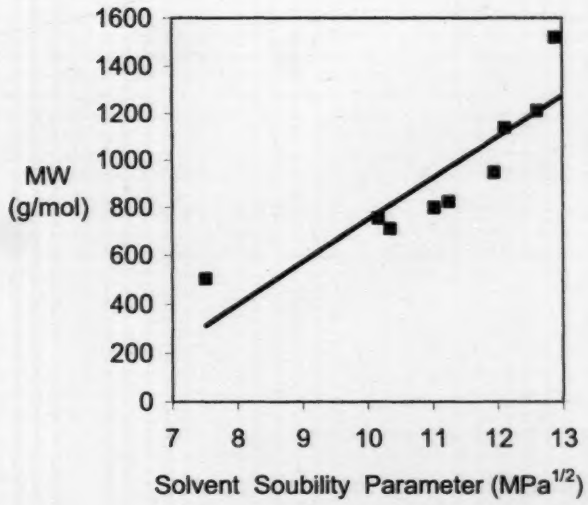


Figure 4-3. Solubility parameters of the ten residue fractions.

Fraction 10 represents the material that was not soluble in supercritical pentane. Since Athabasca asphaltenes are soluble in a heptane/toluene mixture, a solubility parameter of 8.5 was used. The value for the solubility parameter for Fraction 1 has more error since it is the average of the solubility parameter of pentane at atmospheric conditions and the upper pressure limit used to collect the fraction.

In the case of Fractions 1-3, saturate molecules were also included in the molecular representations. The optimizations of these fractions took into account the relative distributions of saturates and aromatics. Two molecules from Fraction 1 are presented in Figure 4-4: one of molecules is a saturate (Figure 4-4a) while the other is an aromatic (Figure 4-4b).

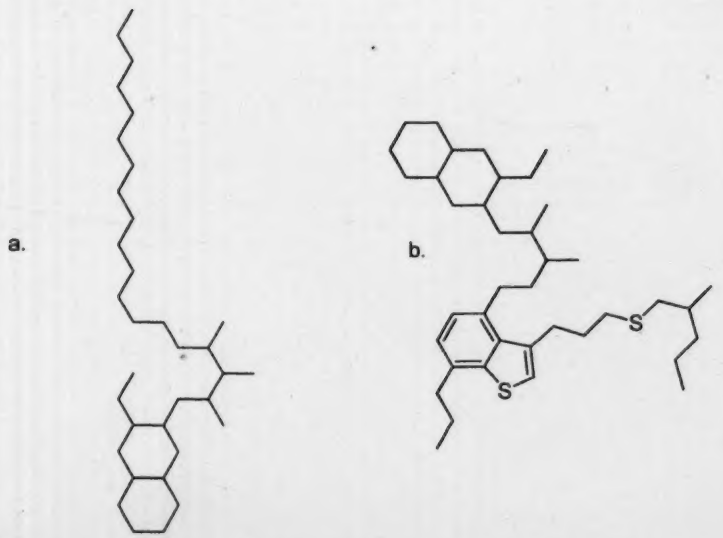


Figure 4-4. Schematic of two molecules from Fraction 1: a. A saturate molecule-$C_{34}H_{66}$; Molecular Weight: 475 g/mol. a. An aromatic molecule-$C_{39}H_{64}S_2$; Molecular Weight: 597 g/mol.

Molecules from Fractions 4, 7 and 10 are presented in Figures 4-5, 4-6, and 4-7. Although the molecular weights of the molecules from the different fractions differ, the general structural frameworks are consistent among fractions: islands of aromatics connected together with aliphatic carbon chains. However, at first glance, although the molecules appear to be very similar, a quantitative approach is required to identify differences between the fractions.

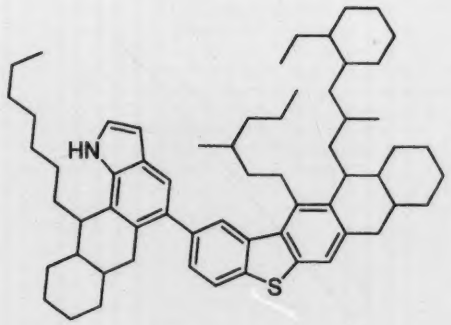


Figure 4-5. Schematic of a molecule from Fraction 4. $C_{62}H_{87}NS$; Molecular Weight: 878 g/mol.

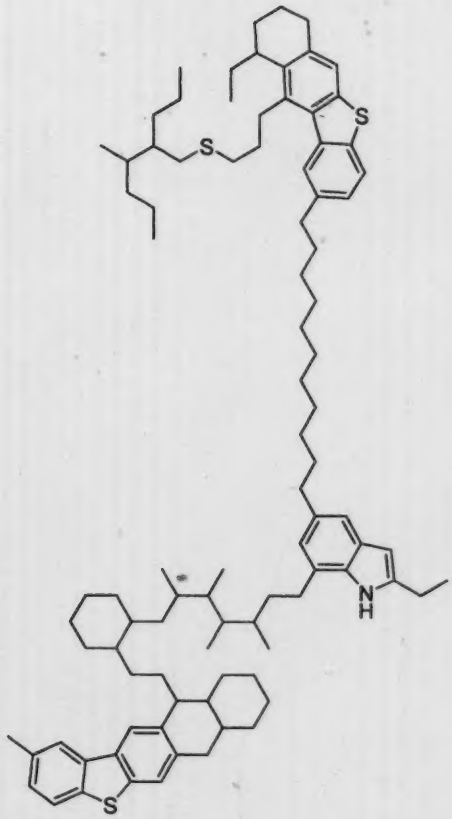


Figure 4-6. Schematic of a molecule from Fraction 7. $C_{92}H_{131}NS_3$; Molecular Weight: 1343 g/mol.

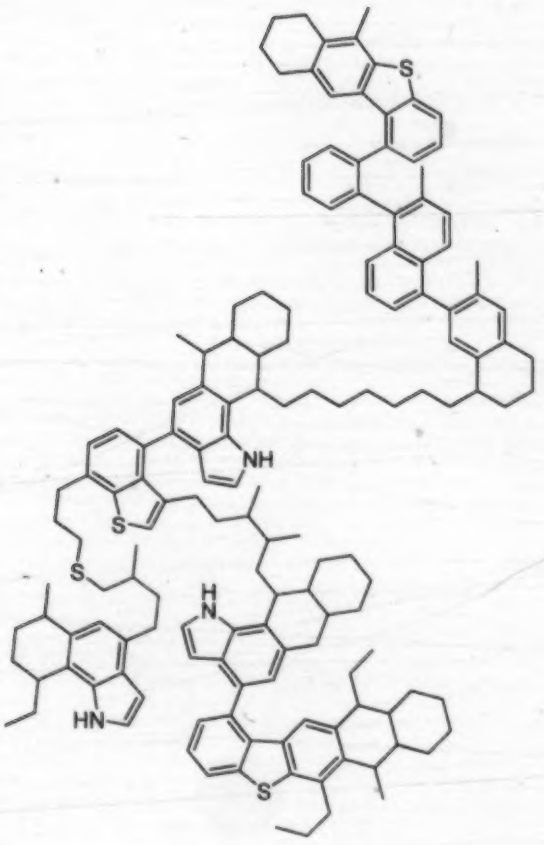


Figure 4-7. Schematic of an asphaltene-like molecule from Fraction 10. $C_{149}H_{171}N_3S_4$; Molecular Weight: 2132 g/mol.

4.3.5. A Consistent Basis

It is important to use a consistent basis when the secondary structure information of one fraction is compared to that of another. In this study, all figures are presented with constant basis was 1000 C. A basis of 100 g was also used, however the trends that were observed were identical to the 1000 C basis. All of the secondary structure information was normalized to reflect the values per 1000 C. If the secondary structure information was not normalized, the absolute value of each property would be skewed by the molecular weight of each fraction.

4.3.6. Aromatic Ring Attachments

Molecular representations allow primary structural information that cannot be calculated directly from analytical techniques such as NMR spectroscopy, to be obtained. Examples of this type of information include the average number of aromatics rings per 1000 carbons in the sample, presented in Figure 4-8, and the average number of aromatic rings per aromatic cluster, presented in Figure 4-9. To present the data, the results from each fraction are plotted against the average solubility parameter for the fraction. The saturate molecules in Fractions 1-3 have a fundamentally different primary structure than the aromatic containing molecules. Thus, Fractions 1-3 have been excluded from Figures 4-8 to 4-12.

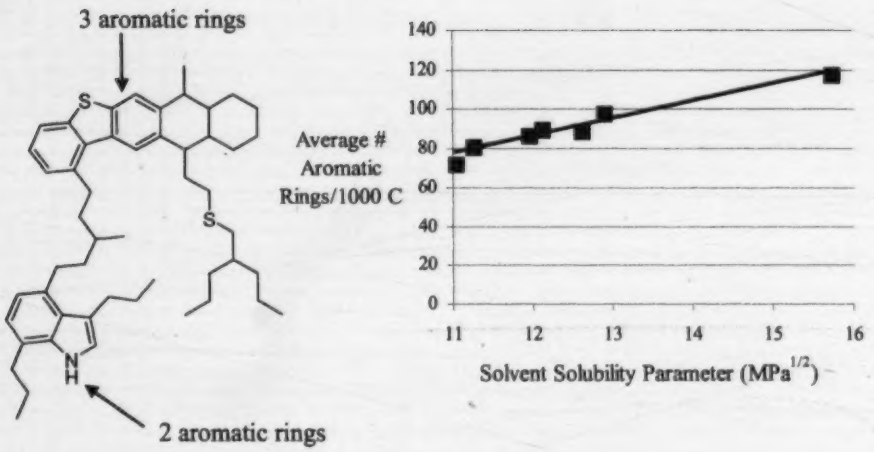


Figure 4-8. Average number of aromatic rings/1000 C. The total number of aromatic rings was calculated for each molecule. For each fraction, the average number of aromatic rings/1000 C was then calculated using the respective mole fractions for each molecule.

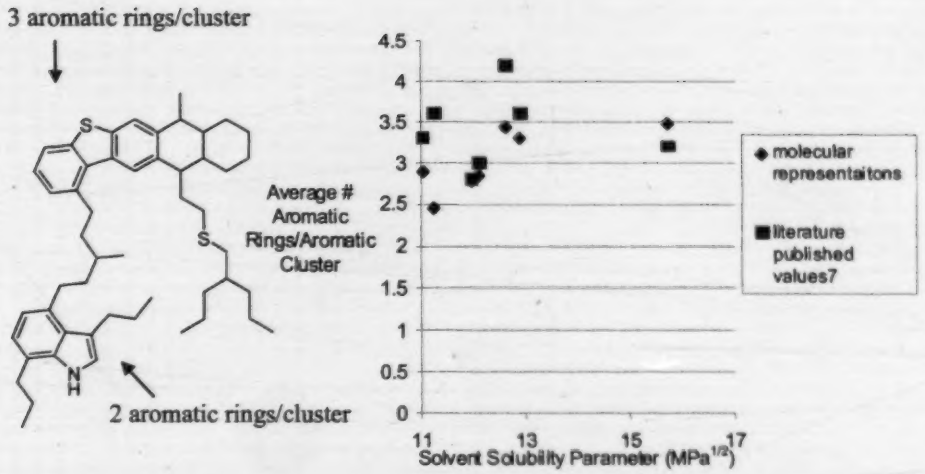


Figure 4-9. Average number of aromatic rings/aromatic cluster. For each molecule in the molecular representation of each fraction, the number of aromatic rings in each aromatic cluster was calculated. The average number of aromatic rings/aromatic cluster for each fraction was then calculated.

The chemical analysis of heavy oil by NMR spectroscopy can yield a wealth of structural information such as aliphatic content and aromaticity. Aromaticity is useful in evaluating changes in the fractions with molecular weight. In Table 4-1, it can be seen that the aromaticity increases with each fraction, and fraction 10, the residue fraction, has the highest aromaticity. In terms of physical property and reactivity modeling, aromaticity is not very useful. A more useful construct for reactivity studies is the number of aromatic rings per 1000 carbons in a fraction. Figure 4-8 shows the increase in the number of aromatic rings per 1000 carbons as a function of the average solubility of the fraction. This information is not obtainable directly from NMR spectroscopic analysis but can be calculated from quantitative molecular representations. In the construction of the molecules, the archipelago framework was employed. In this construction, aromatic cores were connected by aliphatic bridges. The size of the cores was not fixed and was allowed to be varied during the construction and optimization

process. If the pericondensed single core model was assumed, as the aromaticity and number of aromatic rings per 1000 carbons increased, the size of the cores would also be expected to increase. Figure 4-9 presents the average number of aromatic rings per aromatic cluster for each fraction.

Zhao et. al. calculated the number of aromatic rings/cluster for the same bitumen fractions analyzed in this study.[7] However, the analytical data used by Zhao et al. is not the same data used in this thesis. Proton NMR spectroscopy data and the Brown and Ladner method were used by Zhao *et al.* to calculate the average number of aromatic rings/cluster for each fraction. This data is also presented in Figure 4-9. In general, there is very good agreement between the data from the molecular representations, and the fraction data published by Zhao et al. Only fractions 3 and 8 had significant deviations.

George and Beshai used HPLC separation to analyze the relative distirubtion of aromatic ring classes in deasphalted Athabasca bitumen residue.[12] Based upon the data presented in this study, the average number of aromatic rings/cluster for an Athabasca deasphalted bitumen ranges between 2.7 and 3.3 rings/cluster. This is also in good agreement with the data presented in Figure 4-9.

While the aromaticity increased with solubility parameter, the average size of the aromatic cores was relatively constant. This result is interesting in that it suggests that on the basis of size, the structural units of the heavy molecules are relatively similar and that the larger molecules have more of these units connected together. Given the low yield of valuable products from the cracking of asphaltenes, there is more to consider than just aromatic cluster size.

4.3.7. Aromatic Attachments

The average distance between two aromatic cores is another factor that stays relatively constant among the different fractions (Figure 4-10). The primary structural unit in the residue molecules is an aromatic cluster connected to a chain of aliphatic carbon. Thus, the overall aromatic-aliphatic primary structural units of the molecules in the different fractions are very similar from one fraction to another. However, there are noticeable differences in both the abundance and size of the side chains connected to the aromatics in the various fractions. In Figure 4-11, it is apparent that as molecular weight increased, the average length of the side chains connected to the aromatics decreased. Likewise, Figure 4-12 illustrates that as the molecular weight of the fractions increased, the average number of side chains per aromatic cluster decreased. These two observed phenomena could be a result of an actual difference in structural chemistry between the various fractions. Conversely, these observed differences between the frequency and size of aromatic side chains, could either have been caused by constraints used in the construction algorithm or by a bias in the optimization procedure. Molecules with fewer and shorter aromatic side chains have higher aromaticites compared to similar molecules with more and longer side chains. Thus, the aromaticity constraint in the objective function, might have created a bias towards selecting molecules with fewer and shorter aromatic side chains, in order to satisfy the aromaticty constraint. The algorithm that connected the aromatic and aliphatic groups together contained detailed rules that insured that every group was connected to at least one other group. As the size of the molecules increased, it is logical that the connectivity rules would position aliphatic carbon between aromatic clusters, rather than to side chains. During the construction of molecules, if too much aliphatic carbon was assigned to side chains, all of the sampled aromatic groups would have not been connected together.

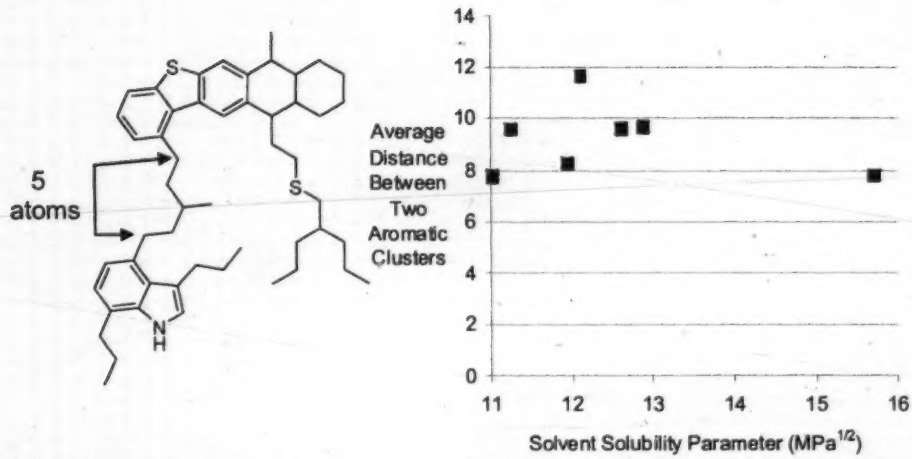


Figure 4-10. Average distance between aromatic clusters. The distances between aromatic clusters were calculated by counting the number of atoms between each pair of adjacent aromatic clusters. For each fraction, the average distance between a pair of aromatic clusters was then calculated.

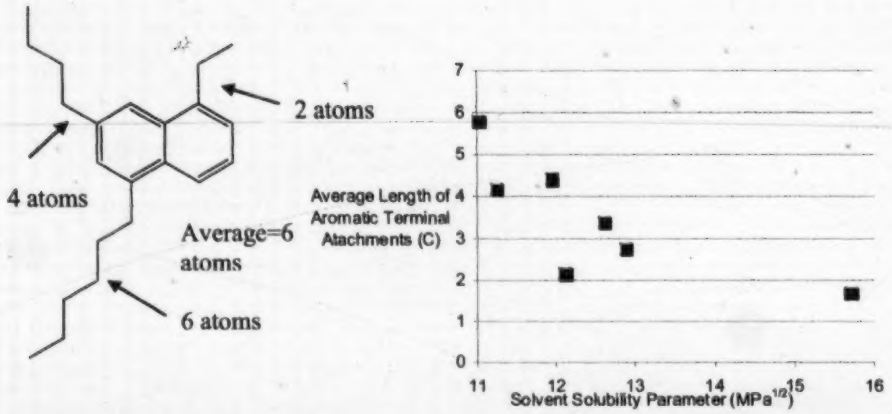


Figure 4-11. Average length of aromatic terminal attachments. The lengths of aromatic side chains were calculated by counting the number of atoms between a terminal carbon and the closest α-attached carbon relative to the closest aromatic ring. For each fraction, the average side chain length was then calculated.

Stausz *et al.* used Ruthenium Ion Catalyzed Oxidation to analyze the distribution of C-C bridges connected between aromatic groups, and the distribution of aromatic C-C side chains in Athabasca asphaltene.[13] Bridges were identified with lengths up to 24 carbons. Based on the results of this study, an average aromatic-aromatic C-C bridge length was about 6.4 carbons. This is consistent with the average aromatic-aromatic C-C calculated for the fraction 10 molecular representation that is presented in Figure 4-10. In the Stausz *et al.* study, aromatic C-C side chains were identified with lengths up to 32 carbons with an average of about 5 carbons. This is higher than the average aromatic C-C side chain length of about 2 carbons calculated in this study for the fraction 10 molecular representation that is presented in Figure 4-11.

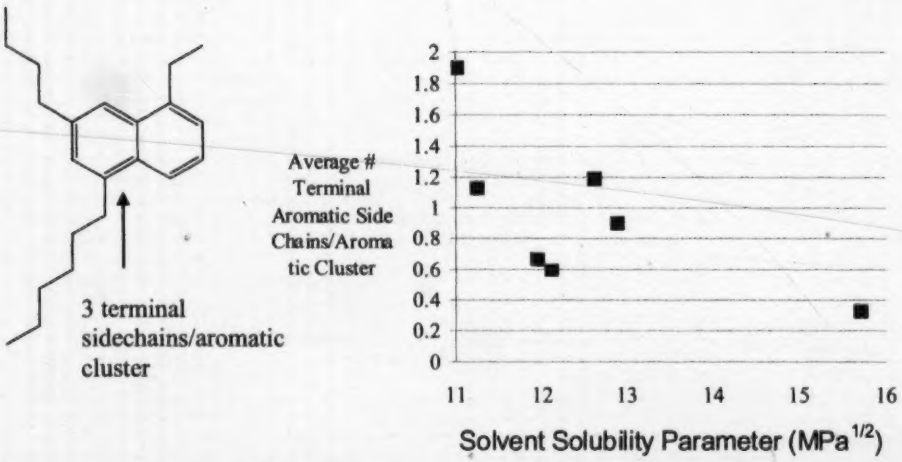


Figure 4-12. Average number of aromatic terminal attachments. The total number of aromatic side chains was calculated for each aromatic cluster in each molecule and the average number of side chains per aromatic cluster was calculated for each fraction.

Molecular representations yield a wealth of data that describe the intermolecular connections of molecules – the primary structure. Specifically, molecular representations describe how the aromatics are connected to the aliphatic C-C chains – both side chains, and aromatic-aromatic linking chains. This information is useful in predicting how molecules will react during thermal processing. The direct NMR spectroscopy data defines the structural groups that are present but it does not describe how the groups are connected together within molecules. Molecular representations can detect and define structural differences between fractions from the same petroleum source, or differences between samples of different petroleums.

This type of analysis is very effective in identifying biases and trends in the primary structures of molecules. However, careful thought, and additional experimentation will be necessary to distinguish between a bias in the

construction or optimization methodology, or an actual occurrence in the structural chemistry of the source material.

4.4. Conclusions

Quantitative molecular representations can be used to generate small sets of molecules that can be used to gain insight into the structure and properties of the heavy oil fractions. The conclusions in this section are based on the data and analysis presented in this chapter. However, good fit does not mean that proposed structures are in fact correct. The molecules in the various super critical fractions vary in molecular weight. On a per carbon basis, the various molecular representations both have properties that differ, and properties that stay relatively constant. As molecular weight increases, the number of aromatic rings per unit weight also increases, but the number of rings per aromatic group is relatively constant in the various molecular representations. Also, as the average molecular representation molecular weight increases, both the number of side chains attached to aromatics and the length of the side chains attached to the aromatic groups decreases. However, the distance between aromatic groups stays relatively constant. Thus, molecular representations provide primary structural information that can not be obtained directly from the raw data.

4.5. References

1. Bartle, K.D. Martin, T.G. Williams, D.F. *Fuel.* **1975**, *54*, 226-235.

2. Neurock, M.; Nigam, A.; Trauth, D.; Klein, M.T. *Chem. Eng. Sci.* **1994**, *49*, 4153-4177.

3. Campbell, D M.; Klein, M.T. *Applied Cat. A: General* **1997**, *160*, 41-54.

4. Redelius, P.G. *Fuel.* **2000**, *79*, 27-35.

5. Mannistu, K. D.; Yarranton, H. W.; Masliyah, J. H. *Energy & Fuels* **1997**, *11*, 615-622.

6. Zhao, S, Sparks, B.D., Kotlyor, L.B., Chung, K.H. *Pet. Sci. Tech.* **2002**, *20(9)*, 1071-1087.

7. Zhao, S.; Kotlyar, L.S.; Woods, J.R.; Sparks, B.D.; Gao, J.; Chung, K.H. *Pet. Sci. Technol.* **2003**, (1-2), 183-199.

8. Chung, K.H.; Xu, C.M.; Hu, X.Y.; Wang, R.N. *Oil Gas J.* **1997**, *95(1)*, 66-70.

9. Yarrangton, H. W.; Alboudwarej, H.; Jakher, R. *Ind. Eng. Chem. Res.* **2000**, *39(8)*, 2916-2924.

10. Yen, T.F. *Prepr. Div. Petrol. Chem. Soc.* **1972**, *17(4)*, F102-F110.

11. Strausz, O.P.; Mojelsky, T.W.; Faraji, F.; Lown, E.M. *Energy Fuels* **1999**, *13(2)*, 207-227.

12. George, A.E.; Beshai, J.E. *Fuel.* **1983**, *62(3)*, 345-349.

13. Strausz, O.P.; Mojelsky, T.W.; Lown, E.M. *Fuel* **1992**, *71(12)*. 1355-1363.

Chapter 5. Thermal Cracking Simulations of Bitumen Molecular Representations

5.1. Introduction

The objective of the research presented in this chapter was to model the thermal cracking of bitumen using molecular representations and the thermal reaction data from previously published model compound studies. Data was obtained for the thermal cracking of Athabasca bitumen under mild conditions without catalyst. A probabilistic thermal cracking model was developed that used literature data for the cracking of various petroleum model compounds. The molecular representation methodology developed and presented in Chapters 3 and 4 was used to create representations of the bitumen feed. Simulated distillation data was used to create the bitumen feed molecular representations. The cracking model simulated the thermal reactivity of petroleum molecules in mild process conditions. One of the objectives of the cracking simulations was to verify that the molecular representations generated the expected range of thermal cracking products when they are reacted in a thermal cracking simulation. A secondary goal of this work was to compare the thermal cracking simulation performance of molecular representations (*Summa*) created from data for the entire feed, to molecular representations (*Distributio*) created from the data for fractions (kerosene, gas oil, residue, etc.) that compose the feed.

5.2. Data

The reactivity and analytical work documented in this section was completed by NCUT (National Center For Upgrading Technology). The NMR spectroscopic analyses and interpretations in this study were performed by Dr. Heather Dettman at NCUT. An Athabsca bitumen sample was thermally cracked in a 1-liter stirred tank reactor. The stir-tank reactor contained 400 g of oil

sample was first flushed with nitrogen and then heated to 405℃. The temperature was held for 30 minutes and then allowed to cool back to room temperature before the gas and liquid products were collected. The run was performed a second time and the two sets of liquid products were mixed before analyses.

D1160 Distillation was used to separate both the Athabasca feed and liquid products into three fractions: kerosene (IBP to 204℃), gas oil (204 to 524℃), and residue (+524C℃). The relative produc t distributions and concentrations of various chemical groups of the liquid feed and product fractions are presented in Tables 5-1 and 5-2. VPO at 130℃ using *o*-dichlorobenzene was used to determine the average molecular weights of the feed and product fractions. Elemental analysis was used to determine the percent composition of C, H, N, and S for each feed and product fraction. The fractions were analyzed using simulated distillation. Figure 5-1 presents the overall simulated distillation curve for the feed and liquid product, while the simulated distillation curves for the three feed fractions are presented in Figure 5-2.

Table 5-1. Distillation breakdown, elemental analysis, and group distribution of the Athabasca bitumen feed.

Yield		0.60	44.76	54.64	100.00
Sample	OFI (i) *	ABFK*	ABFG*	ABFR*	ABF*
Molecular Weight (g/mol)	1	181	345	1112	549
Elemental Content (% total mass)		Wt %	Wt %	Wt %	Wt %
Carbon	2	87.12	85.09	84.84	84.97
Hydrogen	3	11.45	11.47	8.33	9.76
Nitrogen	4	0.33	0.16	0.48	0.34
Sulphur	5	1.09	3.28	6.34	4.94
Total (% total mass)		100.00	100.00	100.00	100.00
Carbon Type (% total C)		Mole % C	Mole % C	Mole % C	Mole % C
Aromatic	6	10.85	25.64	40.08	33.44
Cycloparaffinic	7	51.23	36.05	28.41	31.97
Branched Paraffin	8	5.85	8.89	2.82	5.55
Paraffin Chain (>C2)	9	3.54	5.42	11.71	8.85
Olefin	10	0.34	0.11	0.26	0.20
Ring M & E & DPM		13.06	11.11	6.75	8.74
Chain attachments		15.14	12.77	9.98	11.26
Total (% total C)		100.00	100.00	100.00	100.00
Aromatic and Olefin (% total C)					
Aromatic NS (C+CH)	11	1.59	3.21	5.14	4.25
**Aromatic Bridge C	12	0.00	3.60	10.98	7.61
Aromatic Alkyl Subst. C	13	6.76	11.25	10.13	10.61
Aromatic CH	14	2.50	7.58	13.83	10.96
Methyl-Aromatic	15	1.84	2.07	1.62	1.82
Ethyl-Aromatic	16	2.53	1.95	1.15	1.52
Propyl-Aromatic	17	0.44	0.62	0.71	0.67
***Diphenyl Methane CH_2	18	0.00	2.63	1.43	1.96
***Olefin CH	19	0.28	0.01	0.06	0.04
****Olefin CH_2	20	0.06	0.11	0.20	0.16
Cycloparaffin and Paraffin (% total C)					
Cycloparaffin Bridge CH	21	5.48	8.16	1.95	4.75
Cycloparaffin Alkyl Subst. CH	22	14.09	8.63	5.57	6.99
Cycloparaffin CH2	23	31.66	19.26	20.88	20.22
Methyl-Cycloparaffin	24	4.12	1.87	0.78	1.29
**Ethyl-Cycloparaffin	25	4.56	2.59	1.77	2.15
Paraffin Chain	26	3.10	4.80	11.00	8.17
Branched Paraffin (% total C)					
**Methyl-Branched Chain	27	3.86	8.39	2.82	5.32
**Ethyl-Branched Chain	28	1.30	0.18	0.00	0.09
+Butyl-Branched Chain	29	0.69	0.32	0.00	0.15
Chain Attachments (% total C)					
α to Sulphides	30	0.00	0.00	1.44	0.79
α or β to Aromatic Rings	31	7.01	5.47	3.92	4.63
α to Cycloparaffin Rings	32	7.69	5.47	3.90	4.63
α or β to Branch Points	33	0.45	1.84	0.72	1.22
Total (% total C)		100.00	100.00	100.00	100.00

*OFI refers to the Objective Function Index, *ABFK refers to the kerosene feed fraction, ABFN refers to the kerosene feed fraction, ABFR refers to the residue feed fraction, ABF refers to the entire feed.

Table 5-2. Distillation breakdown, elemental analysis, and group distribution of the liquid Athabasca bitumen thermal cracking products.

Yield		7.71	56.01	33.68	97.40
Sample	OFI (i) *	ABPK*	ABPG*	ABPR*	Total Liquid Product
Molecular Weight (g/mol)	1	131	315	886	355
Elemental Content (% total mass)		Wt %	Wt %	Wt %	Wt %
Carbon	2	84.56	85.20	85.44	83.02
Hydrogen	3	13.37	11.05	7.55	9.76
Nitrogen	4	0.01	0.20	0.83	0.39
Sulphur	5	2.07	3.55	6.18	4.23
Total (% total mass)		100.00	100.00	100.00	97.40
Carbon Type (% total C)		Mole % C	Mole % C	Mole % C	Mole % C
Aromatic	6	16.12	34.78	55.61	39.46
Cycloparaffinic	7	16.94	25.55	14.29	20.43
Branched Paraffin	8	21.01	10.21	4.73	8.93
Paraffin Chain (>C2)	9	20.90	8.37	9.25	9.41
Olefin	10	2.14	0.31	0.25	0.42
Ring M & E & DPM		17.33	9.02	5.21	8.14
Chain attachments		5.56	11.77	10.65	10.61
Total (% total C)		100.00	100.00	100.00	97.40
Aromatic and Olefin (% total C)					
Aromatic NS (C+CH)	11	1.85	2.83	2.95	2.72
Aromatic Bridge C	12	3.60	8.53	19.22	11.53
Aromatic Alkyl Subst. C	13	6.79	10.69	13.36	11.01
Aromatic CH	14	3.89	12.73	20.09	14.20
Methyl-Aromatic	15	3.40	1.96	1.01	1.70
Ethyl-Aromatic	16	2.48	1.92	1.32	1.71
Propyl-Aromatic	17	2.50	0.80	0.59	0.84
Diphenyl Methane CH_2	18	0.00	1.95	0.98	1.42
Olefin CH	19	1.85	0.27	0.21	0.36
Olefin CH_2	20	0.29	0.04	0.04	0.06
Cycloparaffin and Paraffin (% total C)					
Cycloparaffin Bridge CH	21	0.00	3.66	1.52	2.58
Cycloparaffin Alkyl Subst. CH	22	3.90	6.02	4.01	5.02
Cycloparaffin CH_2	23	13.04	15.87	8.76	12.85
Methyl-Cycloparaffin	24	4.25	0.00	0.00	0.33
Ethyl-Cycloparaffin	25	7.20	3.18	1.92	2.98
Paraffin Chain	26	18.41	7.57	8.66	8.57
Branched Paraffin (% total C)					
Methyl-Branched Chain	27	18.10	9.23	4.07	7.94
Ethyl-Branched Chain	28	1.88	0.85	0.40	0.76
+Butyl-Branched Chain	29	1.03	0.12	0.26	0.24
Chain Attachments (% total C)					
α to Sulphides	30	0.00	0.70	4.13	1.78
α or β to Aromatic Rings	31	2.63	4.40	2.55	3.53
α to Cycloparaffin Rings	32	0.00	4.42	3.05	3.51
α or β to Branch Points	33	2.93	2.25	0.91	1.79
Total (% total C)		100.00	100.00	100.00	97.40

*OFI refers to the Objective Function Index, *ABFN refers to the kerosene product fraction, ABFN refers to the kerosene product fraction, ABFR refers to the residue product fraction, ABF refers to the entire liquid product.

A modified version of the NMR spectroscopy method documented in Chapter three of this thesis was used to analyze the three Athabasca liquid feed and three product fractions. The major improvements of this method over the previous method detailed in Chapter three include:

-the ability to resolve the "other" carbons into α to Aromatic Rings, β to Aromatic Rings, α to Branch Points, β to Branch Points, and α to Cycloparaffin Rings.

-more explicit cycloparaffin carbon assignments

-the assignment of terminal CH_3 carbons to both aromatic and naphthenic ring systems

-the assignment of terminal propyl CH_3 carbons to aromatics

-the assignment of diphenyl methane

-the assignment of terminal ethyl CH_3 carbons to naphthenic ring systems

Example molecules that illustrates the various groups that were quantified using NMR spectroscopy are presented in Figures 5-3 and 5-4.

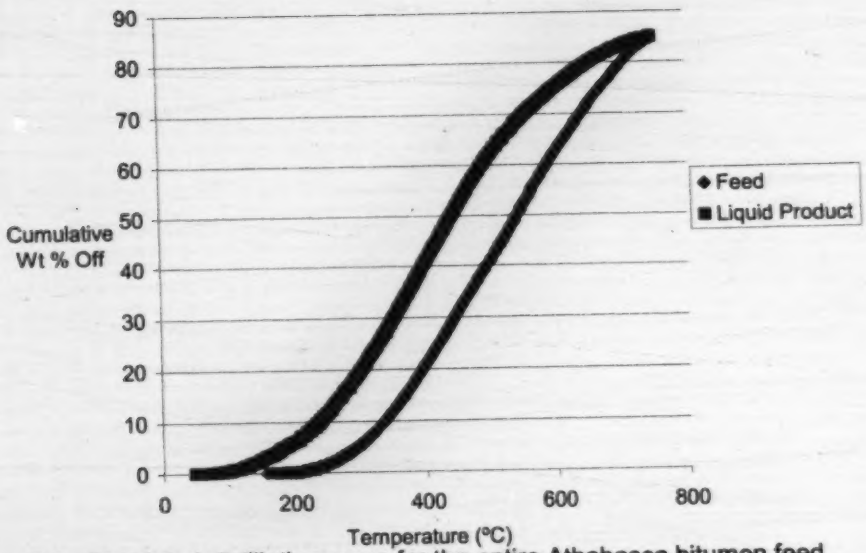


Figure 5-1. Simulated distillation curve for the entire Athabasca bitumen feed and liquid product.

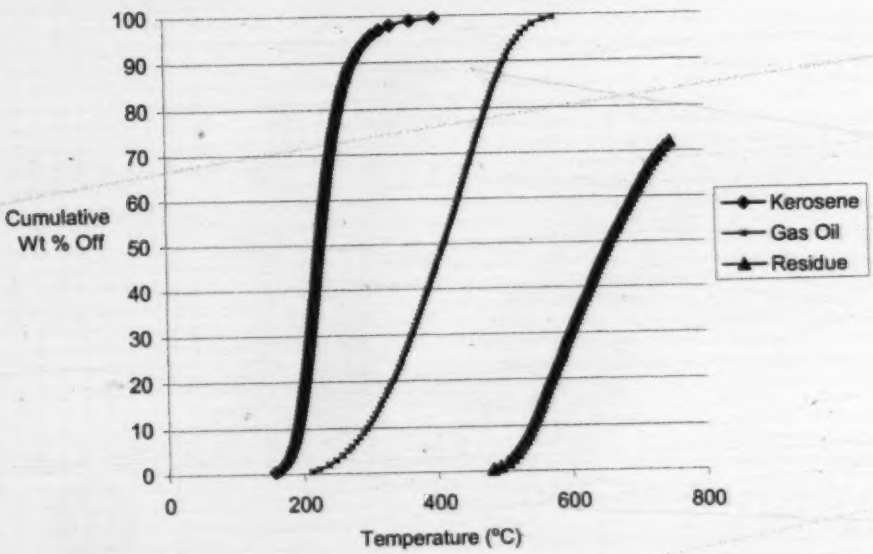


Figure 5-2. Simulated distillation curve for the Athabasca bitumen kerosene, gas oil and residue fractions.

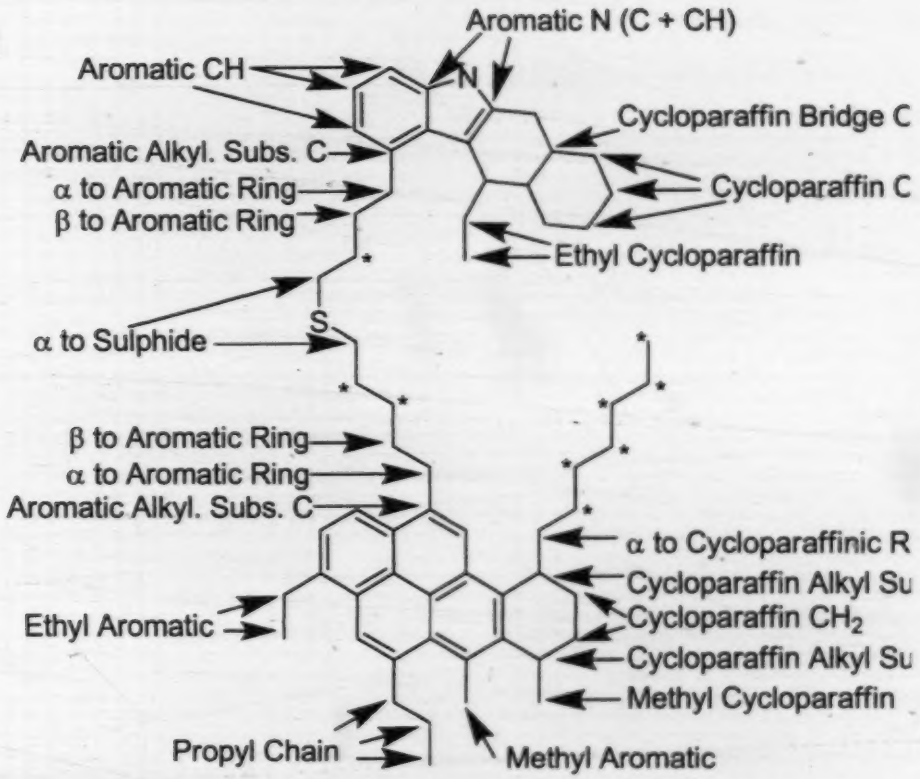


Figure 5-3. Group assignments obtained from the modified NMR spectroscopic heavy oil analysis method.

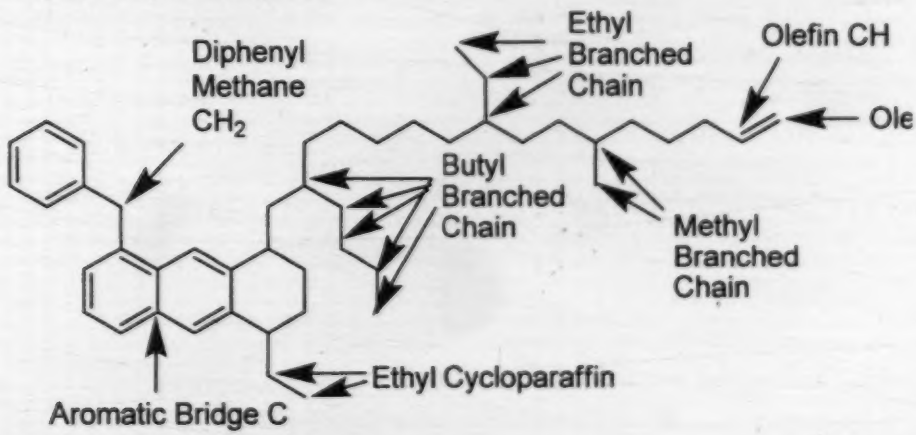


Figure 5-4. Additional group assignments obtained from the modified NMR spectroscopic heavy oil analysis method.

5.3. Thermal Cracking Modeling

The literature model compound thermal cracking studies described in Chapter 2 were used to model the thermal conversion of petroleum molecules. Figures 5-5 to 5-9 illustrate the model compounds used in these studies. The literature model compound studies are discussed in detail in Appendix A. A detailed description of the reaction modeling is presented in Appendix D. The model compound studies discussed in Chapter 2 allow the cracking probabilities of C-C and C-S bonds to be determined by their proximity to key structural groups. These key structural groups featured are: CH branches[1], thioethers[2], aromatic rings[3], pyrene[4] rings, and naphthenic[5] rings. The model compound studies present the relative distributions of reaction products from thermal cracking reactions involving the model compounds. The main objective in this study was to utilize this data to predict the relative distribution of cracked products from thermal cracking simulations.

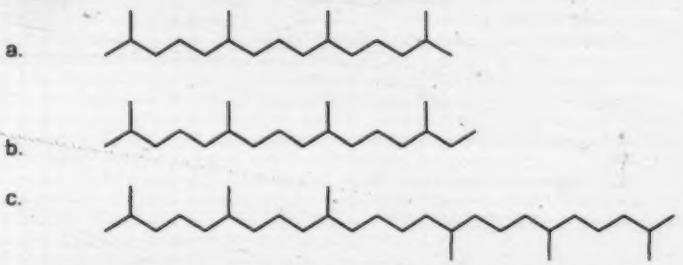


Figure 5-5. The structures of pristane (a.), phytane (b.), and squalane (c.).[1]

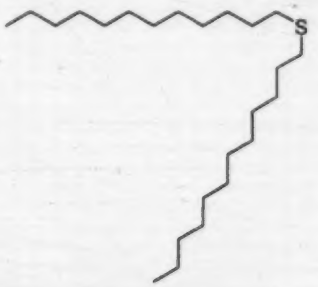


Figure 5-6. The structures of didodecyl sulfide.[2]

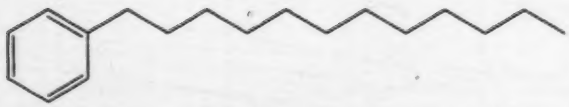


Figure 5-7. The structure of phenyldodecane.[3]

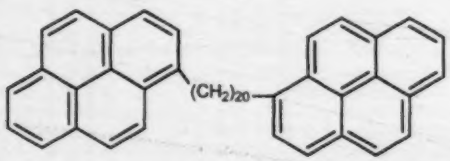


Figure 5-8. The structure of 1,20-di(1-pyrenyl)eicosane.[4]

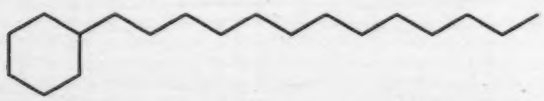


Figure 5-9. The Structures of: a. Tridecylcyclohexane[5].

The model compound data were used to calculate relative cracking propensities (RCPs). A relative cracking propensity (RCP) represents the likelihood of a specific bond – defined by its proximity to a structural group - cracking relative to a bond located a significant distance from a structural group. Thus, the higher a relative cracking propensity of a bond, the greater the probability that the bond will crack during thermal conversion. Figure 5-10 presents the relative cracking propensities calculated from the literature data. Relative cracking propensities are calculated at specific bond distances relative to specific structural group. The bond directly adjacent to a structural group is defined as being "0" bonds away from a structural group. The RCPs presented in Figure 5-10 are normalized relative to the average of reaction products 8, 9 , and 10 carbons away from the structural groups in the reaction studies. The bond with the highest relative cracking propensity was the C-S bond. Other

reactive bonds include: the C-C bond adjacent to a phenyl group, the second C-C bond from a pyrene ring, the C-C bond adjacent to a cyclohexane ring, and the C-C bond adjacent to a CH branched carbon.

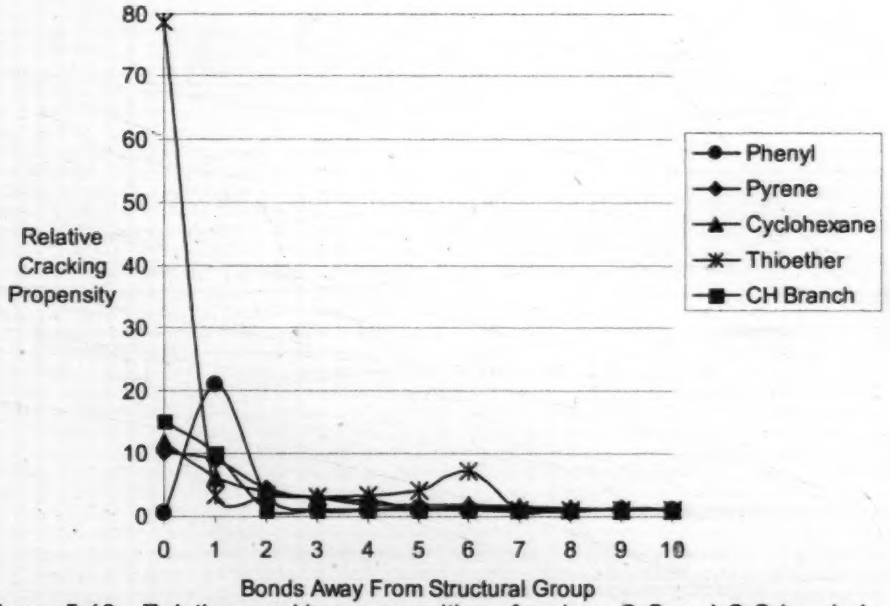


Figure 5-10. Relative cracking propensities of various C-C and C-S bonds in a variety of model compounds. The following respective model compounds and literature references were used to construct this plot: ■CH Branch- Pristane, Phytane, Squalane[1], ∗Thioether- Didodecyl Sulfide[2], •Phenyl – Phenyldodecane[3], ♦Pyrene-1,20-di(1-pyrenyl) Eicosane[4], ▲Cyclohexane-TridecylCyclohexane.[5]

A thermal cracking algorithm was developed that used both the relative cracking propensities and molecular representations as inputs. This algorithm is described in detail in Appendix D. This algorithm used the relative cracking propensities to calculate the probability of breaking each specific bond in the molecular representations. The probability of breaking a specific bond was defined as the relative cracking propensity of the bond divided by the summation of all the RCPs for the molecule Once bond breaking probabilities were calculated, the algorithm stochastically sampled C-C and C-S bonds to crack.

The algorithm would regenerate the connection and structural matrices of each cracked product. The algorithm also reassigned the carbon assignments of each cracked product, and calculated the concentration of each cracked product. After each cracking event, the thermal cracking algorithm determined if the products were gas or liquid. Liquid products were allowed to react further. Gas products were not allowed to react further. The cracking algorithm stochastically reacted ten molecules at each reaction step. Cracking was allowed to continue until the production of IBP-250°C material was consistent with the analytically determined value.

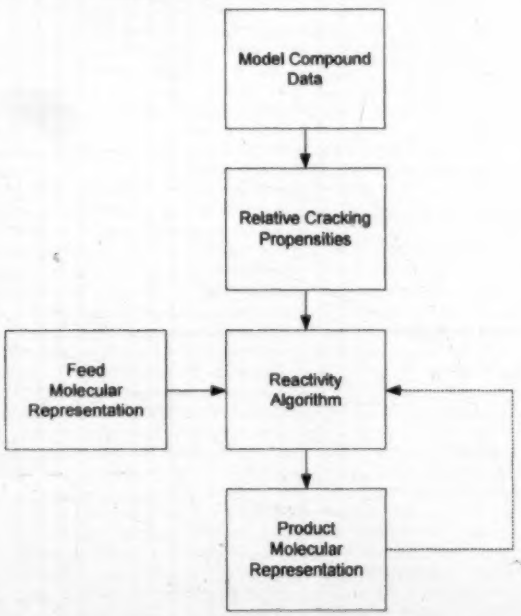


Figure 5-11. Overview of the thermal cracking algorithm.

The reactions modeled in this study are presented in Figure 5-12. Reaction 1 represents the breakage of a C-C bond while reaction 2 represents the breakage of a C-S bond. The relative cracking propensities for these reactions are presented in Figure 5-10. Reaction 3 represents the dehydrogenation of naphthenic rings to aromatic rings. When a molecule was stochastically chosen to crack, its napthenic rings were stochastically sampled to

change to aromatic rings. Each naphthenic ring was given a 50% chance of transformation. Preliminary simulations using various probabilities of naphthenic to aromatic transformation identified 50% as optimal. Only one naphthenic ring in two ring systems was allowed to convert to an aromatic ring.

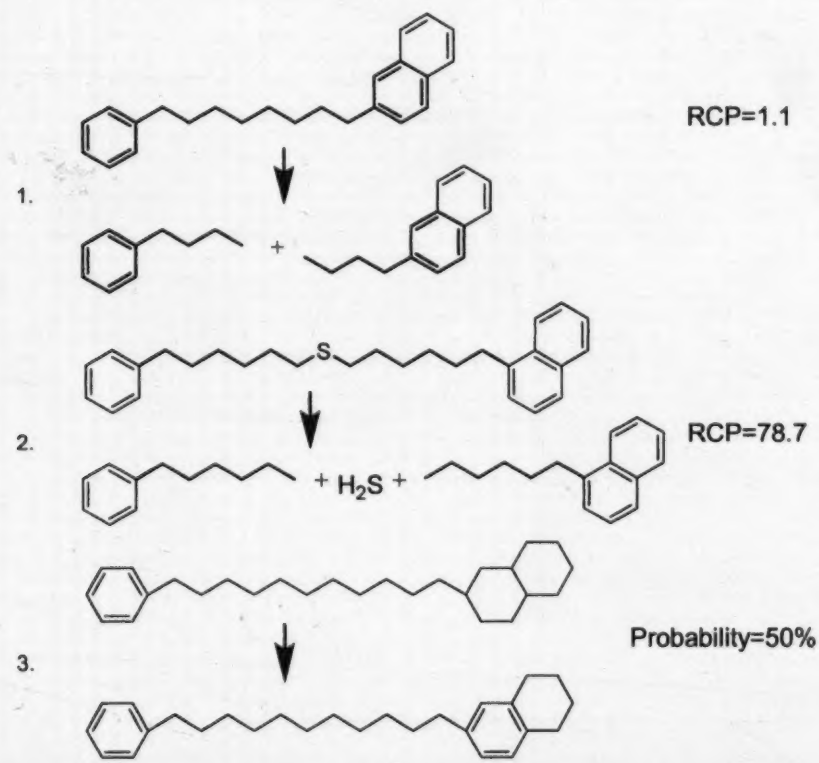


Figure 5-12. Transformations and reactions modeled in the thermal cracking simulations. 1. C-C bond cracking. 2. C-S bond cracking. 3. Dehyrdogenation of naphthenic rings to aromatic rings. RCP represents relative cracking propensity.

5.4. Molecular Representations

5.4.1. Construction Method

The NMR spectroscopy data used in this chapter required a new method of molecular construction to represent all of the groups that it resolved. An overview of the molecule construction method used in this section is presented in Figure 5-13 and a detailed description is found in Appendix E.

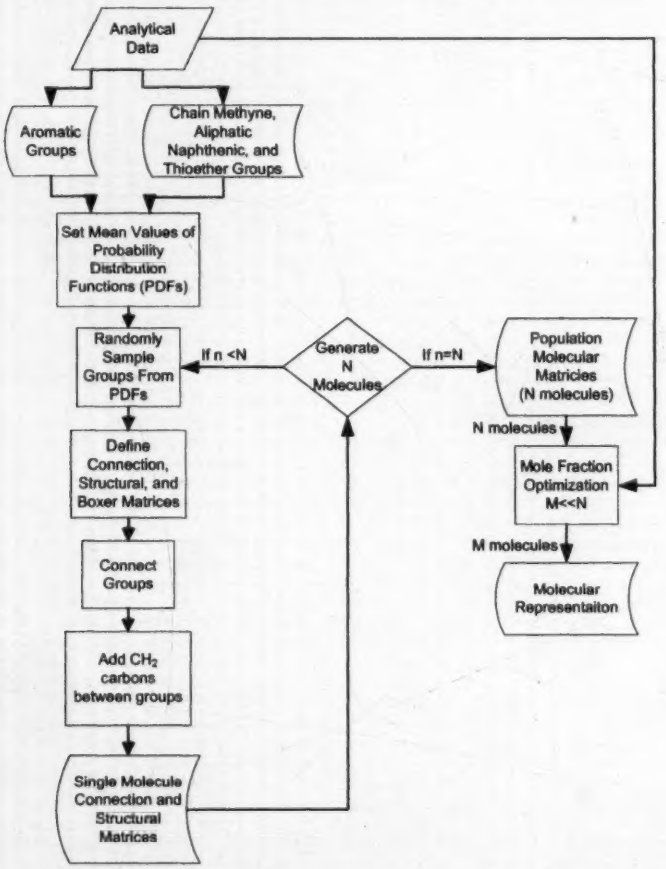


Figure 5-13. Molecular representation construction algorithm flowchart.

5.4.2. Objective Function Formulation

A sum of squares objective function was used for the optimizations (Equations 5-1 and 5-2):

$$F = \sum_{n=1}^{33} \left[\frac{\mu'_n - \mu_n}{\epsilon_n} \right]^2 + D \quad (5\text{-}1)$$

Where:

$$\mu'_n = \sum_{i=1}^{M} x_i \mu'_{n,i} \quad (5\text{-}2)$$

x_i is the mole fraction of each molecule in a set, $\mu'_{n,i}$ is the value of property "n" calculated from the molecular representation, μ_n is the experimentally determined value of property "n" (described in Table 5-1), ε_n is the estimated error of property "n", "*M*" represents the total number of molecules in a molecular representation, and "*D*" takes into account the differences between the experimental simulated distillation data and the boiling point distribution of the molecules in the molecular representation.

$$D = 5 \sum_{i=1}^{12} \left[\frac{A_i - B_i}{A_i} \right]^2 \quad (5\text{-}3)$$

The Marrero-Gani group contribution method was used to calculate boiling points of the molecules in the molecular representation and cracking studies[8]. A_i represents the total percent mass of oil in region "i" in a simulated distillation curve found between T_{iL} and T_{iU}. B_i represents the total percent mass of molecules in a molecular representation found between T_{iL} and T_{iU}.

5.5. Molecular Representation Results

200 molecules were created for each feed fraction: kerosene, middle distillate, and residue. Sequential optimization was then used to produce molecular representations for each fraction that were consistent with the experimental data. For each of the three feed fractions, fifteen molecules were necessary to produce a molecular representation that had good agreement with the experimental data. For the residue fraction, a plot of the objective function versus number of molecules is presented in Figure 5-14. The "D" term in the objective function (Equation 5-3) was only included after three molecules were in the representation. The objective function gradually declined as the number of molecules increased from three to fifteen. Similar trends were observed for the other two fractions.

The molecular representations for the kerosene, gas oil, and residue fractions were then combined to create an overall *Distributio* molecular representation consisting of 45 molecules. The overall feed data was used to construct a *Summa* molecular representation for the entire feed. This *Summa* molecular representation contained 17 molecules.

The data in Tables 5-3 to 5-7 present optimized data for the naptha, gas oil, reside, *Distributio*, and *Summa* molecular representations. In general, there was very good agreement between the predicted properties and the experimental data. For the overall *Summa* molecular representation of the feed, there was very good agreement between the calculated and experimental properties (Table 5-7). The data in Tables 5-8, 5-9, 5-10, and 5-11 present a detailed breakdown of the atomic contents of each molecule in each representation and their respective molecular weights, aromaticites, and boiling points.

Figures 5-15 to 5-19 present comparisons between the experimental simulated distillation data and the boiling distributions calculated from the various

molecular representations. There was excellent agreement between the simulation data and the calculated boiling point distribution for the gas oil, residue, and *Summa* feed molecular representations. The kerosene representation was underrepresented by light components and therefore the calculated boiling point curve deviated from the simulated distillation curve. The boiling point curve for the *Distributio* feed molecular representation closely matched the simulated distillation data for the feed.

Overall, the *Summa* feed representation fit the chemical data better compared to the *Distributio* molecular representation. However, the *Distributio* molecular representation was able to fit the simulation distillation data better than the *Summa* molecular representation.

Examples of various molecules from the kerosene molecular representation are presented in Figure 5-20. Examples of various molecules from the gas oil molecular representation are presented in Figure 5-21. Figures 5-22 to 5-25 present various molecules from the residue molecular representation.

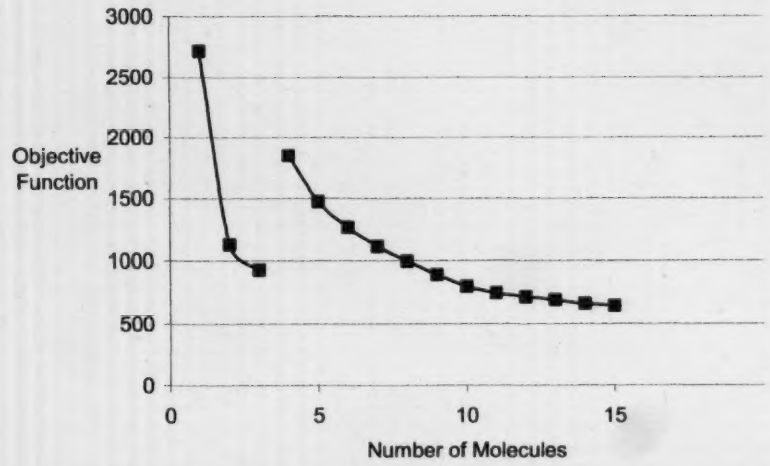


Figure 5-14. A plot of the residue molecular representation objective function as a function of number of molecules included in the molecular representation. The residue molecules were selected from an initial population of 200 molecules.

Table 5-3. Experimental versus calculated properties of the kerosene molecular representation. 15 molecules were present in the final representation. Sequential optimization was used to select 15 molecules from an initial population of 200 molecules.

Property	Error	Experimental (E)	Calculated (C)	(C-E)/E*100%
Molecular Weight (g/mol)		181	194	7.3
Elemental Content (% total mass)				
Carbon	1	87.1	85.6	1.8
Hydrogen	1	11.5	13.0	13.8
Nitrogen	0.2	0.3	0.3	0.1
Sulphur	0.2	1.1	1.1	0.3
Total (% total mass)		100.0	100.0	
Carbon Type (% total C)				
Aromatic	3	10.9	11.4	5.2
Cycloparaffinic	3	51.2	48.7	5.0
Branched Paraffin	3	5.9	18.5	216.6
Paraffin Chain (>C2)	3	3.5	3.8	7.2
Chain attachments	3	15.2	6.5	57.0
Aromatic and Olefin (% total C)				
Aromatic NS (C+CH)	1	1.6	1.6	1.4
Aromatic Bridge C	1	0.0	0.8	
Aromatic Alkyl Subst. C	1	6.8	3.2	52.2
Aromatic CH	1	2.5	5.8	130.5
Methyl-Aromatic	1	1.8	1.2	37.1
Ethyl-Aromatic	1	2.5	2.0	21.5
Propyl-Aromatic	1	0.4	0.5	21.8
Diphenyl Methane CH_2	1	0.0	0.0	0.0
Olefin CH	1	0.3	0.3	17.2
Olefin CH_2	1	0.1	0.1	1.2
Cycloparaffin and Paraffin (% total C)				
Cycloparaffin Bridge CH	1	5.5	7.0	28.5
Cycloparaffin Alkyl Subst. CH	1	14.1	9.2	34.4
Cycloparaffin CH2	1	31.7	32.4	2.2
Methyl-Cycloparaffin	1	4.1	3.9	5.2
Ethyl-Cycloparaffin	1	4.6	3.1	31.6
Paraffin Chain	1	3.1	3.8	22.4
Branched Paraffin (% total C)				
Methyl-Branched Chain	1	3.9	15.6	303.4
Ethyl-Branched Chain	1	1.3	2.1	0.0
+Butyl-Branched Chain	1	0.7	0.9	0.0
Chain Attachments (% total C)				
α to Sulphides	1	0.0	0.0	0.0
α or β to Aromatic Rings	1	7.0	1.5	78.5
α to Cycloparaffin Rings	1	7.7	4.9	36.8
α or β to Branch Points	1	0.5	0.2	66.6
Total		100	100	1206.8

Table 5-4. Experimental versus calculated properties of the gas oil molecular representation. 15 molecules were present in the final representation. Sequential optimization was used to select 15 molecules from an initial population of 200 molecules.

Property	Error	Experimental (E)	Calculated (C)	(C-E)/E*100%
Molecular Weight (g/mol)	35	346	346	0.1
Elemental Content (% total mass)				
Carbon	1	85.1	85.3	0.3
Hydrogen	1	11.5	11.3	1.8
Nitrogen	0.2	0.2	0.2	0.0
Sulphur	0.2	3.3	3.3	0.4
Total (% total mass)		100	100	
Carbon Type (% total C)				
Aromatic	3	25.6	23.6	7.9
Cycloparaffinic	3	36.1	41.1	13.9
Branched Paraffin	3	8.9	12.5	40.6
Paraffin Chain (>C2)	3	5.4	4.5	17.9
Chain attachments	3	12.8	9.4	26.2
Aromatic and Olefin (% total C)				
Aromatic NS (C+CH)	1	3.2	3.2	0.6
Aromatic Bridge C	1	3.6	1.9	46.7
Aromatic Alkyl Subst. C	1	11.3	9.9	12.1
Aromatic CH	1	7.6	8.6	13.6
Methyl-Aromatic	1	2.1	2.1	2.2
Ethyl-Aromatic	1	2.0	2.1	6.4
Propyl-Aromatic	1	0.6	0.7	14.5
Diphenyl Methane CH_2	1	2.6	0.0	100.0
Olefin CH	1	0.0	0.1	532.0
Olefin CH_2	1	0.1	0.1	42.5
Cycloparaffin and Paraffin (% total C)				
Cycloparaffin Bridge CH	1	8.2	7.5	8.4
Cycloparaffin Alkyl Subst. CH	1	8.6	8.1	5.7
Cycloparaffin CH2	1	19.3	25.5	32.2
Methyl-Cycloparaffin	1	1.9	1.8	1.2
Ethyl-Cycloparaffin	1	2.6	2.1	20.4
Paraffin Chain	1	4.8	4.5	7.3
Branched Paraffin (% total C)				
Methyl-Branched Chain	1	8.4	12.0	43.0
Ethyl-Branched Chain	1	0.2	0.2	0.0
+Butyl-Branched Chain	1	0.3	0.3	0.0
Chain Attachments (% total C)				
α to Sulphides	1	0.0	0.0	0.0
α or β to Aromatic Rings	1	5.5	3.8	31.4
α to Cycloparaffin Rings	1	5.5	4.2	22.9
α or β to Branch Points	1	1.8	1.5	20.7
Total		100	100	1072.9

Table 5-5. Experimental versus calculated properties of the residue molecular representation. 15 molecules were present in the final representation. Sequential optimization was used to select 15 molecules from an initial population of 200 molecules.

Property	Error	Experimental (E)	Calculated (C)	(C-E)/E*100%
Molecular Weight (g/mol)	100	1112	1016	8.6
Elemental Content (% total mass)				
Carbon	1	84.8	83.9	1.1
Hydrogen	1	8.3	9.3	11.2
Nitrogen	0.2	0.5	0.5	0.1
Sulphur	0.2	6.3	6.3	0.5
Total (% total mass)		100	100	
Carbon Type (% total C)				
Aromatic	3	40.1	40.6	1.4
Cycloparaffinic	3	28.4	27.1	4.6
Branched Paraffin	3	2.8	3.0	7.5
Paraffin Chain (>C2)	3	11.7	8.2	30.3
Chain attachments	3	10.0	12.7	27.6
Aromatic and Olefin (% total C)				
Aromatic NS (C+CH)	1	5.1	4.2	18.8
Aromatic Bridge C	1	11.0	13.4	22.0
Aromatic Alkyl Subst. C	1	10.1	10.0	1.0
Aromatic CH	1	13.8	13.0	5.7
Methyl-Aromatic	1	1.6	2.0	24.4
Ethyl-Aromatic	1	1.2	1.8	54.1
Propyl-Aromatic	1	0.7	0.5	23.0
Diphenyl Methane CH_2	1	1.4	1.2	18.1
Olefin CH	1	0.1	0.2	150.3
Olefin CH_2	1	0.2	0.2	24.9
Cycloparaffin and Paraffin (% total C)				
Cycloparaffin Bridge CH	1	2.0	4.2	117.8
Cycloparaffin Alkyl Subst. CH	1	5.6	6.4	15.0
Cycloparaffin CH2	1	20.9	16.5	21.2
Methyl-Cycloparaffin	1	0.8	0.3	67.7
Ethyl-Cycloparaffin	1	1.8	2.3	28.4
Paraffin Chain	1	11.0	8.2	25.8
Branched Paraffin (% total C)				
Methyl-Branched Chain	1	2.8	3.0	7.5
Ethyl-Branched Chain	1	0.0	0.0	0.0
+Butyl-Branched Chain	1	0.0	0.0	0.0
Chain Attachments (% total C)				
α to Sulphides	1	1.4	2.4	69.3
α or β to Aromatic Rings	1	3.9	5.5	40.9
α to Cycloparaffin Rings	1	3.9	4.1	5.0
α or β to Branch Points	1	0.7	0.7	6.1
Total		100	100	839.9

Table 5-6. Experimental versus calculated properties of the *Distributio* feed molecular representation. The kerosene, gas oil, and residue molecular representations were combined to create a representation of the entire feed. 45 different molecules represented the feed.

Property	Error	Experimental (E)	Calculated (C)	(C-E)/E*100%
Molecular Weight (g/mol)	50	550	524	4.7
Elemental Content (% total mass)				
Carbon	1	85	84.6	0.5
Hydrogen	1	9.8	10.2	4.1
Nitrogen	0.2	0.3	0.3	0.0
Sulphur	0.2	4.9	4.8	2.0
Total (% total mass)		100	100	
Carbon Type (% total C)				
Aromatic	3	33.4	32.4	3.2
Cycloparaffinic	3	32.0	33.9	5.9
Branched Paraffin	3	5.6	7.6	37.3
Paraffin Chain (>C2)	3	8.9	6.4	27.9
Chain attachments	3	11.3	11.1	1.3
Aromatic and Olefin (% total C)				
Aromatic NS (C+CH)	1	4.3	3.7	13.1
Aromatic Bridge C	1	7.6	7.9	3.5
Aromatic Alkyl Subst. C	1	10.6	9.9	6.6
Aromatic CH	1	11.0	10.9	0.6
Methyl-Aromatic	1	1.8	2.1	13.0
Ethyl-Aromatic	1	1.5	1.9	26.1
Propyl-Aromatic	1	0.7	0.6	6.9
Diphenyl Methane CH_2	1	2.0	0.6	0.0
Olefin CH	1	0.0	0.1	175.5
Olefin CH_2	1	0.2	0.1	32.3
Cycloparaffin and Paraffin (% total C)				
Cycloparaffin Bridge CH	1	4.8	5.8	22.0
Cycloparaffin Alkyl Subst. CH	1	7.0	7.2	3.6
Cycloparaffin CH2	1	20.2	20.8	3.0
Methyl-Cycloparaffin	1	1.3	1.0	20.0
Ethyl-Cycloparaffin	1	2.2	2.2	1.3
Paraffin Chain	1	8.2	6.4	21.9
Branched Paraffin (% total C)				
Methyl-Branched Chain	1	5.3	7.4	38.4
Ethyl-Branched Chain	1	0.1	0.1	0.0
+Butyl-Branched Chain	1	0.2	0.2	0.0
Chain Attachments (% total C)				
α to Sulphides	1	0.8	1.3	0.0
α or β to Aromatic Rings	1	4.6	4.7	0.6
α to Cycloparaffin Rings	1	4.6	4.2	10.2
α or β to Branch Points	1	1.2	1.0	14.5
Total		100	100	500.2

Table 5-7. Experimental versus calculated properties of the *Summa* feed molecular representation. 17 molecules were present in the final representation. Sequential optimization was used to select 17 molecules from an initial population of 200 molecules

Property	Error	Experimental (E)	Calculated (C)	(C-E)/E*100%
Molecular Weight (g/mol)	50	524	550	8.2
Elemental Content (% total mass)				
Carbon	1	85	84.5	0.6
Hydrogen	1	9.8	10.3	5.1
Nitrogen	0.2	0.3	0.3	0.0
Sulphur	0.2	4.9	4.9	0.0
Total (% total mass)		100	100	
Carbon Type (% total C)				
Aromatic	3	33.4	32.0	4.6
Cycloparaffinic	3	32.0	33.2	3.8
Branched Paraffin	3	5.6	6.8	17.9
Paraffin Chain (>C2)	3	8.9	7.5	18.6
Chain attachments	3	11.3	12.5	9.6
Aromatic and Olefin (% total C)				
Aromatic NS (C+CH)	1	4.3	3.8	12.6
Aromatic Bridge C	1	7.6	9.4	18.9
Aromatic Alkyl Subst. C	1	10.6	8.9	19.6
Aromatic CH	1	11.0	9.9	10.4
Methyl-Aromatic	1	1.8	2.0	10.2
Ethyl-Aromatic	1	1.5	1.5	1.3
Propyl-Aromatic	1	0.7	0.7	0.7
Diphenyl Methane CH_2	1	2.0	0.7	0.0
Olefin CH	1	0.0	0.1	62.5
Olefin CH_2	1	0.2	0.1	49.8
Cycloparaffin and Paraffin (% total C)				
Cycloparaffin Bridge CH	1	4.8	5.4	11.6
Cycloparaffin Alkyl Subst. CH	1	7.0	6.7	3.8
Cycloparaffin CH2	1	20.2	21.1	4.2
Methyl-Cycloparaffin	1	1.3	1.2	12.0
Ethyl-Cycloparaffin	1	2.2	1.9	12.8
Paraffin Chain	1	8.2	7.5	9.5
Branched Paraffin (% total C)				
Methyl-Branched Chain	1	5.3	6.5	18.3
Ethyl-Branched Chain	1	0.1	0.1	0.0
+Butyl-Branched Chain	1	0.2	0.2	0.0
Chain Attachments (% total C)				
α to Sulphides	1	0.8	1.2	0.0
α or β to Aromatic Rings	1	4.6	5.4	14.3
α to Cycloparaffin Rings	1	4.6	4.7	1.6
α or β to Branch Points	1	1.2	1.1	6.5
Total		100	100	347.8

Table 5-8. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca bitumen kerosene molecular representation[a].

Molecule No.	Mole %	C	H	S	N	Molecular Weight (g/mol)	Aromaticity (% C)	Boiling Point (°C)
1	0.4	15	20	0	0	200	40.0	273
2	1.8	24	36	0	0	325	25.0	376
3	6.2	8	6	1	0	134	100.0	229
4	1.6	14	28	0	0	196	0.0	240
5	20.1	14	26	0	0	194	0.0	251
6	6.3	16	31	0	0	223	0.0	275
7	4.6	11	13	0	1	159	72.7	275
8	2.1	12	18	0	0	162	50.0	226
9	0.4	47	80	1	0	677	17.0	531
10	11.1	12	25	0	0	169	0.0	210
11	18.3	11	22	0	0	154	0.0	187
12	10.5	17	33	0	0	237	0.0	291
13	0.005	36	49	1	1	528	44.4	505
14	7.1	18	30	0	0	246	33.3	310
15	9.6	15	28	0	0	208	0.0	265

[a] 200 molecules formed the initial population. Fifteen molecules were in the final molecular representation.

Table 5-9. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca bitumen gas oil molecular representation[a].

Molecule No.	Mole %	C	H	S	N	Molecular Weight (g/mol)	Aromaticity (% C)	Boiling Point (°C)
1	12.4	17	32	0	0	236	0.0	286
2	0.03	36	51	0	1	498	38.9	489
3	5.8	31	48	1	0	453	25.8	451
4	8.5	22	30	1	0	327	36.4	393
5	3.9	37	51	0	1	510	37.8	508
6	1.6	27	52	0	0	377	0.0	386
7	6.5	21	40	0	0	293	0.0	326
8	1.6	24	36	0	0	325	25.0	375
9	11.4	30	48	0	0	409	20.0	429
10	12.4	22	34	0	0	299	27.3	365
11	2.2	42	68	1	0	605	19.0	510
12	8.6	28	42	1	0	411	28.6	437
13	10.2	33	50	1	0	479	24.2	478
14	6.8	21	32	0	0	285	28.6	347
15	8.2	11	16	0	0	148	54.5	210

[a] 200 molecules formed the initial population. Fifteen molecules were in the final molecular representation.

Table 5-10. Elemental compositions, molecular weights, aromaticities, and boiling points of the fifteen molecules in the Athabasca bitumen residue molecular representation[a].

Molecule No.	Mole %	C	H	S	N	Molecular Weight (g/mol)	Aromaticity (% C)	Boiling Point (°C)
1	2.3	194	272	8	2	2889	35.1	879
2	12.8	45	64	1	0	637	35.6	567
3	4.9	81	134	2	0	1172	9.9	659
4	15.6	37	48	1	0	525	43.2	537
5	1.8	93	109	3	1	1337	47.3	738
6	10.7	83	97	3	1	1205	53.0	712
7	3.5	35	66	1	0	519	0.0	461
8	6.4	106	126	5	1	1575	52.8	767
9	11.6	66	116	0	0	910	0.0	593
10	7.6	64	82	1	0	883	43.8	640
11	3.6	171	211	5	1	2441	46.8	863
12	4.8	59	49	2	1	836	81.4	653
13	7.1	57	66	1	0	783	49.1	626
14	3.1	118	115	5	1	1708	67.8	808
15	4.3	78	100	3	0	1134	46.2	688

Table 5-11. Elemental compositions, molecular weights, aromaticities, and boiling points of the seventeen molecules in the Athabasca bitumen *Summa* feed representation[a].

Molecule No.	Mole %	C	H	S	N	Molecular Weight (g/mol)	Aromaticity (% C)	Boiling Point (°C)
1	13.9	16	22	1	0	246	50.0	331
2	7.9	48	80	1	0	689	16.7	534
3	5.8	69	75	2	1	983	63.8	677
4	3.0	114	121	3	1	1601	59.6	783
5	8.6	16	30	0	0	222	0.0	275
6	9.1	24	45	1	0	366	0.0	396
7	0.0	92	158	3	0	1360	8.7	693
8	0.6	215	283	4	1	3010	37.2	908
9	1.2	28	42	1	0	411	28.6	435
10	1.3	161	173	3	1	2218	63.4	859
11	5.8	11	16	0	0	148	54.5	207
12	6.3	65	105	1	0	919	24.6	630
13	11.9	31	46	1	0	451	25.8	466
14	12.1	32	56	0	0	441	0.0	438
15	6.1	61	108	1	0	874	0.0	589
16	3.2	84	73	1	1	246	78.6	737
17	3.2	23	36	0	0	689	26.1	362

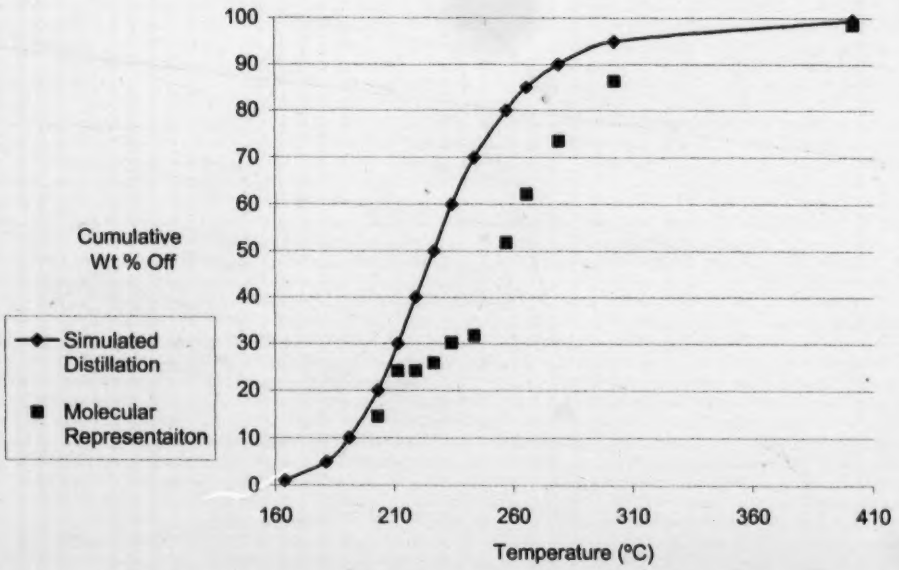


Figure 5-15. Boiling point distribution of the Athabasca bitumen kerosene molecular representation plotted against experimental simulated distillation data.

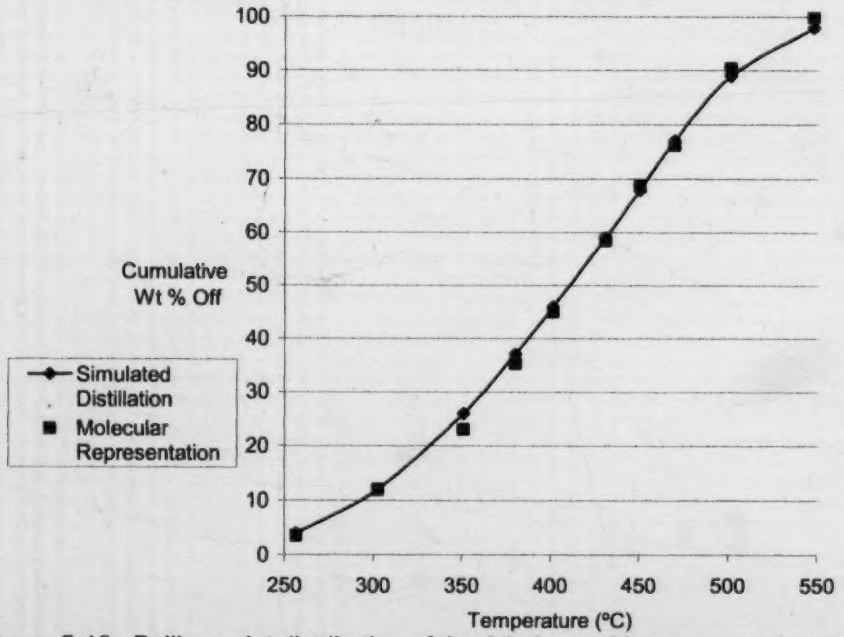


Figure 5-16. Boiling point distribution of the Athabasca bitumen gas oil molecular representation plotted against experimental simulated distillation data.

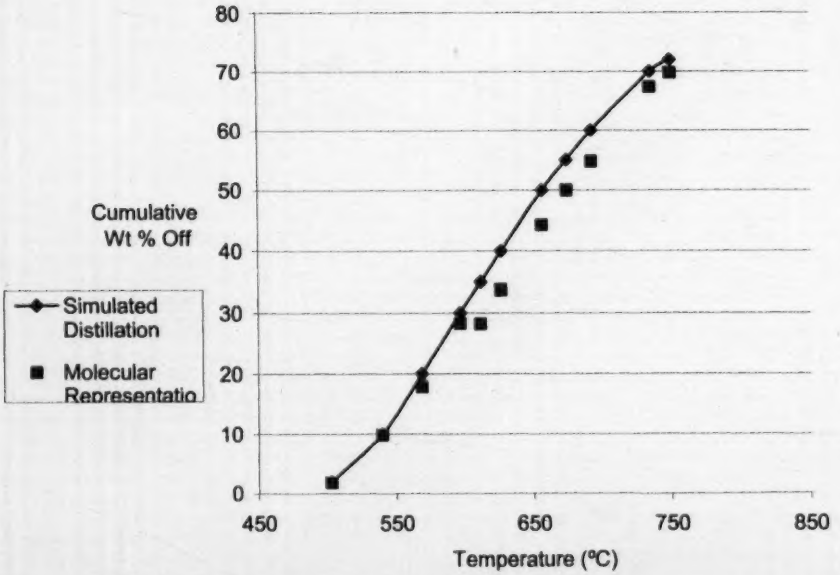


Figure 5-17. Boiling point distribution of the Athabasca bitumen residue molecular representation plotted against experimental simulated distillation data.

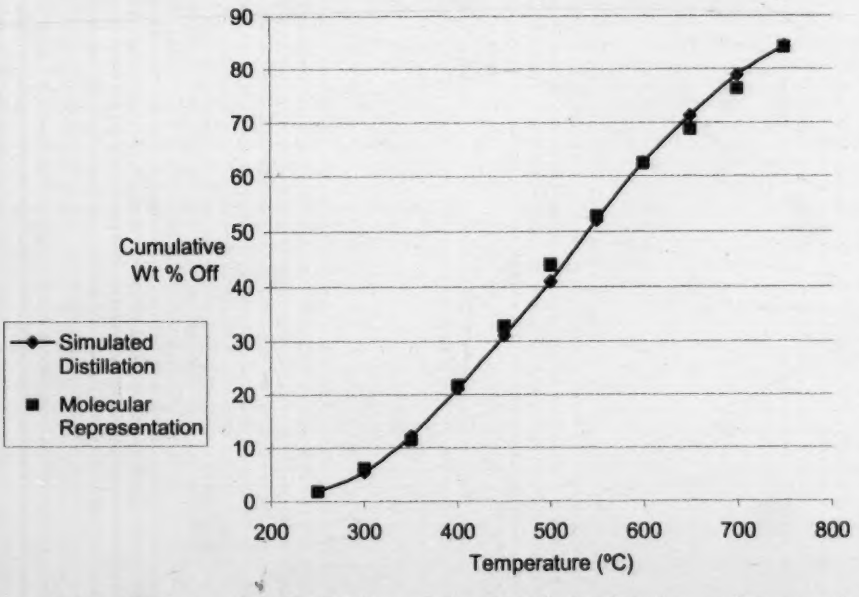


Figure 5-18. Boiling point distribution of the *Distributio* Athabasca bitumen feed molecular representation plotted against experimental simulated distillation data of the entire feed.

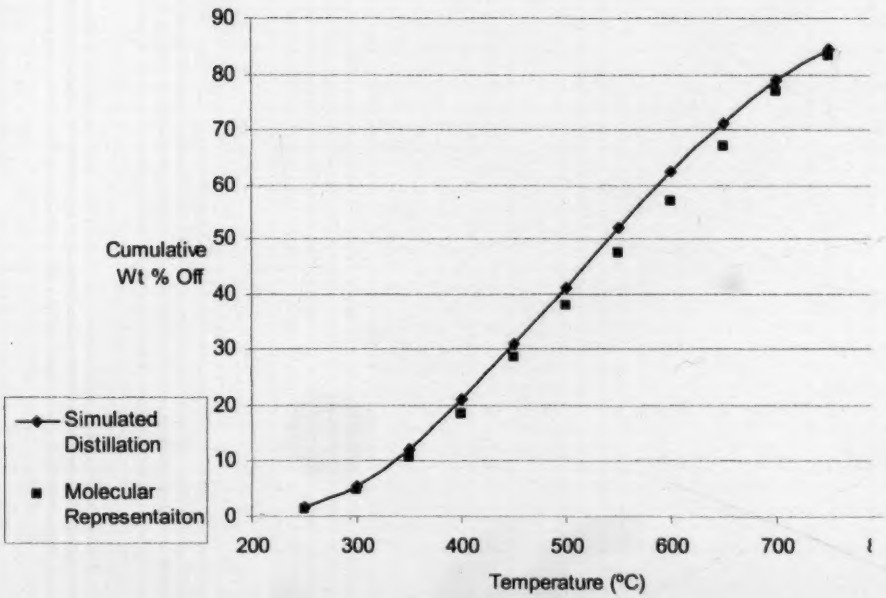


Figure 5-19. Boiling point distribution of the *Summa* Athabasca bitumen feed molecular representation plotted against experimental simulated distillation data of the entire feed.

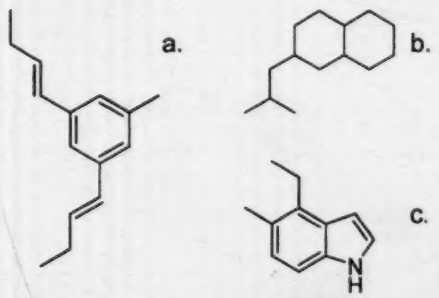


Figure 5-20. Three molecules from the Athabasca bitumen kerosene molecular representation. a. $C_{15}H_{20}$ (Molecular Weight 200.3 g/mol). b. $C_{14}H_{26}$ (Molecular Weight. 194.4 g/mol). c. $C_{11}H_{13}N$ (Molecular Weight 159.2 g/mol).

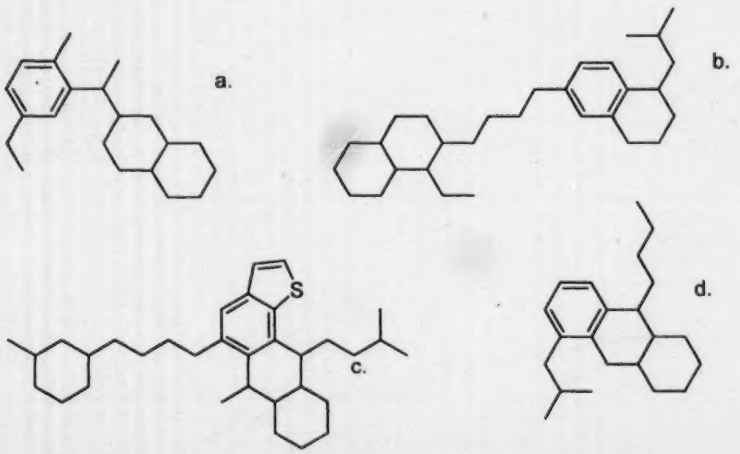


Figure 5-21. Three molecules from the Athabasca bitumen gas oil molecular representation. a. $C_{21}H_{32}$ (Molecular Weight 285 g/mol). b. $C_{30}H_{48}$ (Molecular Weight. 409 g/mol). c. $C_{33}H_{50}S$ (Molecular Weight 479 g/mol). d. $C_{22}H_{34}$ (Molecular Weight 299 g/mol).

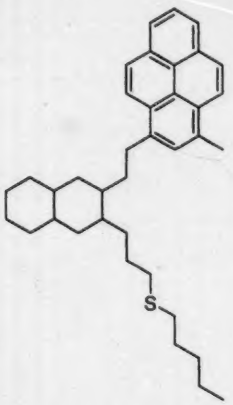


Figure 5-22. Athabasca bitumen residue molecular representation molecule (a.) $C_{37}H_{48}S$ (Molecular Weight 525 g/mol).

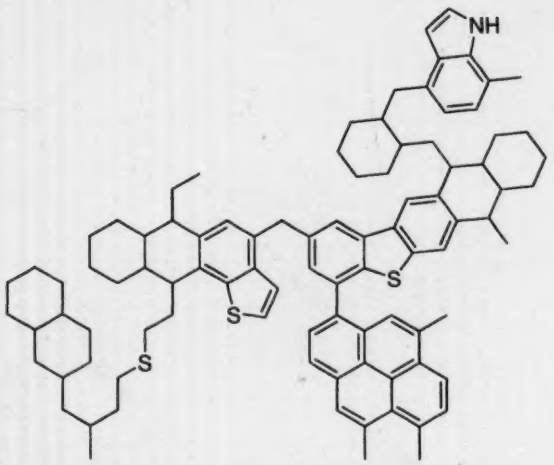


Figure 5-23. Athabasca bitumen residue molecular representation molecule (b.) $C_{93}H_{109}NS_3$ (Molecular Weight 1337 g/mol).

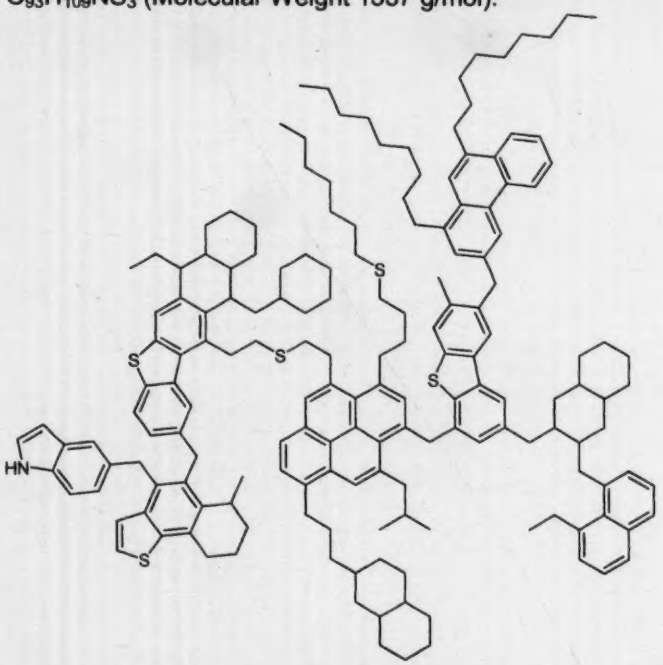


Figure 5-24. Athabasca bitumen residue molecular representation molecule (c.) $C_{171}H_{211}NS_5$ (Molecular Weight 2441 g/mol).

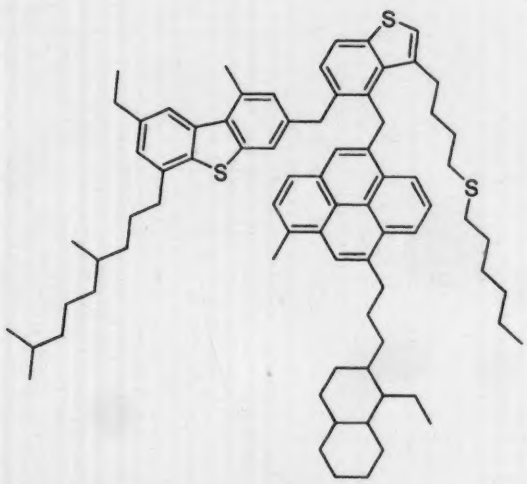


Figure 5-25. Athabasca bitumen residue molecular representation molecule (d.). $C_{78}H_{100}S_3$ (Molecular Weight 1134 g/mol).

5.6. Thermal Cracking Modeling

5.6.1. Model Validation

Molecular representations were created for each model compound in Figures 5-6 to 5-9 (a branched chain was used to study the reactivity of CH branches). In independent simulations, the cracking algorithm was used to crack each molecule. For each model compound simulation, the conversion level was set at one. The number of cracks per molecule was set to be the carbon chain length associated with each model compound. In the validation step, the product molecules were not allowed to react. For each model compound simulation, relative cracking propensities were calculated using the product concentrations. The relative cracking propensities calculated using the model compound cracking simulations were identical to those presented in Figure 5-11. This verification step, assured that the cracking algorithm was bug free and unbiased. It also assured that the cracking products contained the correct carbon type assignments.

5.6.2. The First Thermal Cracking Simulation Method

The first thermal cracking model did not contain the reaction loop in Figure 5-12. It was reasoned that the cracking severity might be so low that once a molecule is cracked, it is very unlikely that the reaction products will react again. Thus, all reaction products were initially not allowed to further react. This restriction significantly reduced the required programming of the cracking simulation. These first simulations, also allowed the mechanics and code to be tested and verified – if they could not perform in a sequential flow, they could not possibly perform in a more complicated loop.

The first thermal cracking simulation used the *Distributio* molecular representation and was characterized by a relatively low degree of conversion of 500°C + material and a low production of IBP- 250°C material (Table 5-12). However, the degree of sulphur removal, formation of intermediate 250-500°C material, and the extent of gas production, were all consistent with the experimental values. The low degree of conversion of 500°C + material combined with the low degree of IBP- 250°C production, rendered it necessary to change the thermal cracking algorithm to allow further reaction of products.

Table 5-12. Results from the thermal cracking simulation that used the *Distributio* feed molecular representation (45 molecules).

Molecular Conversion	Feed	Product	Simulated
% feed in H_2S	*	0.6	0.8
% feed in gas fraction	*	2.6	3.6
# of molecules in product	*	*	222
% <250°C (%mass)*	1.9	12	4.7
% 250-500°C (%mass)*	42.1	51.4	49.3
% 500+°C (%mass)*	56	36.5	46
MW (g/mol) *	524	*	414
% C* (% mass)	84.6	85.2	85.9
% H* (% mass)	10.2	10.0	9.7
% N* (% mass)	0.3	0.4	0.3
% S* (% mass)	4.8	4.3	4.1
% C -aromatics* (% C)	32.4	40.5	44.9
% C -naphthenics* (% C)	33.9	21.0	21.9
% C -branches * (% C)	7.6	9.2	6.3
% C -paraffinics * (% C)	6.4	9.7	6.1
H_2S in Gas (% gas)	*	29.1	20.9
C_1 in Gas (% gas)	*	19.6	7.5
C_2 in Gas (% gas)	*	16.9	2.6
C_3 in Gas (% gas)	*	17.5	8.7
C_4 in Gas (% gas)	*	10.7	29.4
C_5 in Gas (% gas)	*	4.1	18.6
C_6 in Gas (% gas)	*	2.0	1.9
C_7 in Gas (% gas)	*	0.0	3.1
C_8 in Gas (% gas)	*	0.0	1.4
C_9 in Gas (% gas)	*	0.0	5.9

5.6.3. The Second Thermal Cracking Simulation Method

The thermal cracking simulation algorithm was updated to allow a continuous reaction loop. This algorithm allowed the material in the feed molecular representations to be continuously reacted until it could not be reacted any further. In the second method to model the thermal cracking of the molecular representations, integer numbers of each molecule were used, instead of mole fractions. There was very close agreement between the nondescritized molecular representations that used mole fractions and the integer utilizing descretized molecular representations (Tables 5-13 and 5-14).

Both the *Summa* and *Distributio* molecular representations were cracked in separate cracking simulations with a reaction loop. The simulations were divided into reaction steps. At each reaction step, ten molecules were stochastically sampled to crack. The naphthenic rings in the sampled molecules were also sampled (50% probability) to dehydrogenate to aromatic rings. The objective of the cracking simulations was to produce liquid product fractions with IBP- 250°C compositions consistent with the experimentally determined value. The *Distributio* simulations were stopped after 30 reaction steps (Table 5-13) The *Distributio* thermal cracking simulation with the reaction loop (Table 5-13) had a liquid product distribution that was much closer to the experimental data than the *Distributio* cracking simulation without the reaction loop (Table 5-12). In the case of the *Summa* simulation, 40 reaction steps were taken (Table 5-14 and Figures 5-26 to 5-30). The data presented in Tables 5-13 and 5-14 are averages over three independent simulations. However, the molecular analyses presented in Tables 5-15 and 5-16, and Figures 5-26 to 5-33 are for individual unique simulations. For the *Distributio* molecular representation simulation, the percent content of the 500°C+ material in the liquid product fraction was consistent with the experimental data, while the percent content of IBP- 250°C material in the

liquid product fraction deviated about 4 % from the experimental data (Table 5-13 and Figure 5-28). However, the percent content of 250-500℃ material had an approximate 7% deviation relative to the experimental data (Table 5-13). Likewise, more gas (7.3 % feed in gas fraction) was produced in this simulation relative to what was experimentally determined (2.6 % feed in gas fraction) (Table 5-13). In the *Distributio* cracking simulation, the aromaticity increased from 33.2 to 40.5 % (Table 5-13 and Figure 5-30). This was very consistent with the experimentally determined increase from 33.4 to 40.5 %. Likewise, the naphthenic content decreased from 33.0 to 28.6% (Table 5-13 and Figure 5-30). Experimentally, the naphthenic content decreased from 32.0 to 21.0%. The degree of desulphurization in the simulation was also consistent with the experimental data (Table 5-13 and Figure 5-31). In the *Distributio* thermal cracking simulation, the sulfur content decreased from 5.0 to 4.3 weight %. Experimentally, it decreased from 4.9 to 4.3% weight %. The mole average molecular weight of the cracked liquid products decreased from 533 to 338 g/mol in the *Distributio* thermal cracking simulation (Table 5-13 and Figure 5-29). This final liquid product average molecular weight of 338 g/mol was in good agreement with the experimentally determined molecular weight of the cracked liquid products of 355 g/mol.

For the *Summa* molecular representation, in the thermal cracking simulation with the reaction loop, the degree of conversion of the 500°C + material, the production of 250-500°C material, and the production of IBP- 250°C material, were all very consistent with the experimental data (Tables 5-14, and Figures 5-26 to 5-28). The conversion of 500°C + material during a thermal cracking simulation of the *Summa* molecular representation is plotted in Figure 5-26. 40 reaction steps were necessary in this simulation. Slightly more than 50% of the 500°C + material in the feed was converted to product. Figure 5-27 plots the gas production and contents of IBP- 250°C, 250-500°C, and 500°C + material during this same simulation. Similar to the simulation with the *Distributio* molecular representation, more gas (9.7 % feed in gas fraction) was produced in

this simulation relative to what was experimentally determined (2.6 % feed in gas fraction) (Table 5-14). In the *Summa* thermal cracking simulation, the aromaticity increased from 32.5 to 44.3 % (Table 5-14 and Figure 5-30). This was very consistent with the experimentally determined increase from 33.0 to 40.5 %. Likewise, the naphthenic content decreased from 33.0 to 28.6% (Table 5-14 and Figure 5-30). Experimentally, the naphthenic content decreased from 32.8 to 27.0%. The degree of desulphurization in the simulation was also consistent with the experimental data (Table 5-14 and Figure 5-31). In the simulation, the sulfur content decreased from 4.8 to 4.1 weight % (Table 5-14). Experimentally, it decreased from 5.0 to 4.3% weight %. In the *Summa* thermal cracking simulation, the molecular weight decreased from 587 to 326 g/mol (Table 5-14 and Figure 5-29). This value is also in good agreement with the experimentally determined molecular weight of the cracked liquid products of 355 g/mol.

A plot of the experimental liquid product simulated distillation data versus the cumulative boiling point distribution of a liquid *Summa* thermal cracking simulation product is presented in Figure 5-32. There was very good agreement between the experimental and the simulated data.

Table 5-15 presents a summary of the molecular weight, sulphur, aromaticity, and naphthenic values, of both the *Summa* and *Distributio* molecular representation cracking simulations, when the simulated contents of the 500℃+ material matched the experimental liquid product data. The trends displayed here are virtually identical as those stated in the previous paragraphs.

To generate the liquid product molecules in the *Summa* thermal cracking simulation, most feed molecules had to be cracked multiple times (in different reaction steps). Up to eight bonds cracked in a single feed molecule in separate reaction steps (Figure 5-33). At the end of the thermal cracking simulation, the majority of the starting molecules had been cracked, and the majority of the product molecules were the result of two or more cracks of a starting feed

molecule. By comparing the product distribution in Figure 5-27 to the cracking event distribution in Figure 5-33, it is apparent that as the total number of cracks increased, the fit to the experimental data improved as well.

A mole balance of a *Summa* thermal cracking simulation simulation is presented in Table 5-16. On a mole basis, the number of simulated liquid product molecules was consistent with the experimental data. On a mole basis, almost twice as much gas was produced in the simulations relative to what was observed experimentally. Likewise, using the mole balance presented in Table 5-16, about 75% more cracking reactions took place in the thermal cracking simulations relative to the experimental data.

Table 5-13. Gas production, molecular weight, liquid fraction composition, elemental analysis, and carbon type distribution of the descritized *Distributio* molecular representation and the simulated thermal cracking product.

	Nondescritezed Feed	Descritized Feed	Simulation Pròduct	Experimental Product
# molecules in liquid fraction	45	278	403	-
% feed in gas fraction	0	0	7.3	2.6
% feed in H_2S	0	0	1.1	0.6
MW (g/mol) *	524	533	338	355
<250℃ (% mass)*	1.9	2.1	8.0	12
250-300℃ (% mass)*	4.3	4.3	11.8	8.5
300-350℃ (% mass)*	5.3	5.0	12.3	9.9
350-400℃ (% mass)*	10.3	9.8	10.8	11.5
400-450℃ (% mass)*	11.1	10.2	12.7	11.5
450-500℃ (% mass)*	11.1	10.9	10.6	10
500-550℃ (% mass)*	8.7	9.1	6.5	7.7
550-600℃ (% mass)*	9.7	8.8	0.9	5.8
600-650℃ (% mass)*	6.3	6.2	3.2	4.9
650-700℃ (% mass)*	7.6	7.7	12.2	3.5
700-750℃ (% mass)*	7.9	8.3	6.2	2.1
>750℃ (% mass)*	15.8	17.6	4.8	12.5
% <250°C (% mass)*	1.9	2.1	8.0	12
% 250-500°C (% mass)*	42.1	40.2	58.2	51.4
% 500+°C (% mass)*	56	57.7	33.8	36.5
% C (% mass)*	84.6	84.5	85.7	85.2
% H (% mass)*	10.2	10.2	9.8	10
% S (% mass)*	4.8	5.0	4.1	4.3
aromatics (% C) *	32.4	33.2	41.8	40.5
naphthenics (% C) *	33.9	33.0	28.6	21.0
branches (% C) *	7.6	6.4	3.3	9.2

* refers to the physical properties of the liquid fraction

Table 5-14. Gas production, molecular weight, liquid fraction composition, elemental analysis, and carbon type distribution of the descritized *Summa* molecular representation and its simulated cracked product.

	Nondescritized Feed	Descritized Feed	Simulation Product	Experimental Product
# molecules in liquid fraction	17	249	399	-
% feed in gas fraction	0	0	9.7	2.6
% feed in H_2S	0	0	0.9	0.6
MW (g/mol) *	571	587	326	355
<250℃ (% mass)*	1.5	1.5	11.9	12
250-300℃ (% mass)*	3.3	3.2	11.1	8.5
300-350℃ (% mass)*	6	5.7	11.8	9.9
350-400℃ (% mass)*	7.5	7.2	10.2	11.5
400-450℃ (% mass)*	10.2	9.9	5.5	11.5
450-500℃ (% mass)*	9.4	9.0	14.6	10
500-550℃ (% mass)*	9.5	9.0	2.8	7.7
550-600℃ (% mass)*	9.4	9.0	3.5	5.8
600-650℃ (% mass)*	10.1	10.1	4.6	4.9
650-700℃ (% mass)*	9.9	10.3	4.2	3.5
700-750℃ (% mass)*	6.3	6.2	3.3	2.1
>750℃ (% mass)*	16.6	19.0	16.4	12.5
% <250°C (% mass)*	1.5	1.5	11.9	12
% 250-500°C (% mass)*	36.4	35.1	53.2	51.4
% 500+°C (% mass)*	61.8	63.6	34.9	36.5
% C (% mass)*	85	84.5	85.8	85.2
% H (% mass)*	10.1	10.3	9.7	10
% S (% mass)*	4.9	4.8	4.1	4.3
aromatics (% C) *	33.4	32.5	44.3	40.5
naphthenics (% C) *	32	32.8	27.0	21
branches (% C) *	5.6	7.5	3.9	9.2

*refers to the physical properties of the liquid fraction

Table 5-15. Molecular weight, sulfur content, aromaticity, and naphthenic values, of both the *Summa* and *Distributio* molecular representation thermal cracking simulations, when the simulated liquid contents of the 500℃+ material matched the experimental liquid product data

Property	*Summa* Liquid Product (39 Reaction Steps)	*Distributio* Liquid Product (24 Reaction Steps)	Experimental Liquid Product
>500℃ (% liquid mass)	36.4	36.9	36.5
<250℃ (% liquid mass)	11.3	6	12
MW (g/mol)	333	366	355
S (% liquid mass)	4.2	4	4.2
aromatics (% C liquid fraction)	43.7	42.4	39.5
naphthenics (% C liquid fraction)	27.3	27.4	20.4

Table 5-16. Mole balance of a *Summa* molecular representation thermal cracking simulation.

		Avg. Mol Weight (g/mol)	# of Molecules	% Weight of Feed	Mol % of Total Molecules	CH_4 In Gas (Mass % Gas)	CH_4 In Gas (% Feed)	# of Cracking Reactions[2]
Simulated	Liquid Feed	587	249	100	100	*	*	388
	Liquid Product	337	399	91.9	62.6	*	*	
	Gas Product	50	238	8.1	37.4	7.7	0.63	
Experimental[1]	Liquid Feed	550	249	100	100.0	*	*	233
	Liquid Product	355	376	97.4	79.6	*	*	
	Gas Product	37	96	2.6	20.4	18	0.47	

[1]The number of calculated molecules in the experimental liquid fraction was normalized to be consistent with the *Summa* molecular representation.
[2]The number of cracking reactions that occurred is determined by the difference between the number of moles of product (gas and liquid molecules) and feed (liquid).

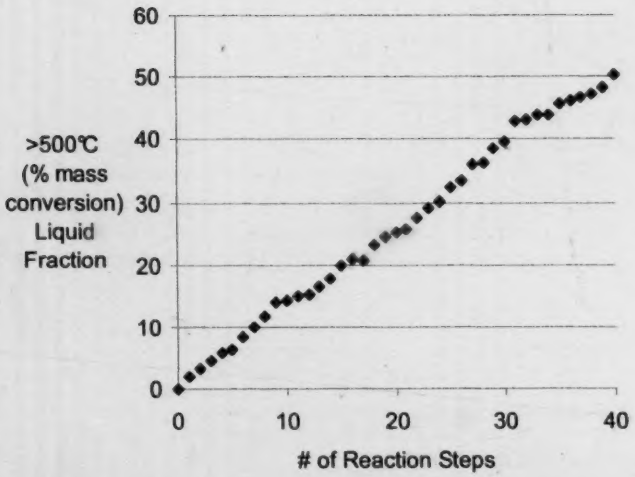


Figure 5-26. Conversion of the 500℃ + material during a thermal cracking simulation that used the *Summa* molecular representation.

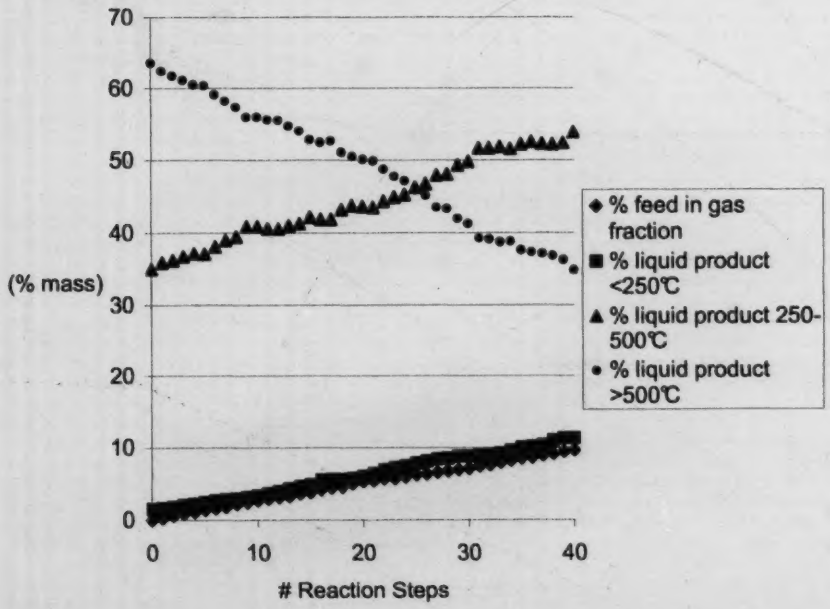


Figure 5-27. Gas production and liquid fraction composition observed during a thermal cracking simulation that used the *Summa* molecular representation.

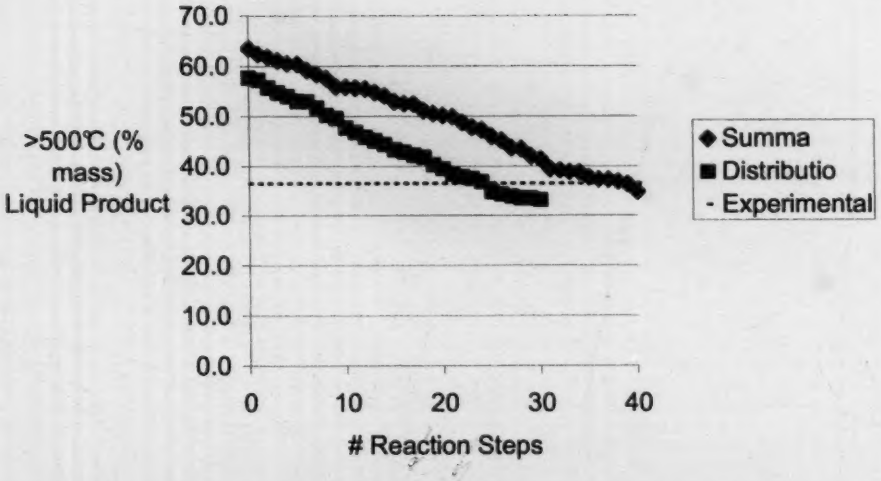


Figure 5-28. Changes in the liquid fraction mass content of 500℃ material during a thermal cracking simulations that used the *Summa* and *Distributio* molecular representations.

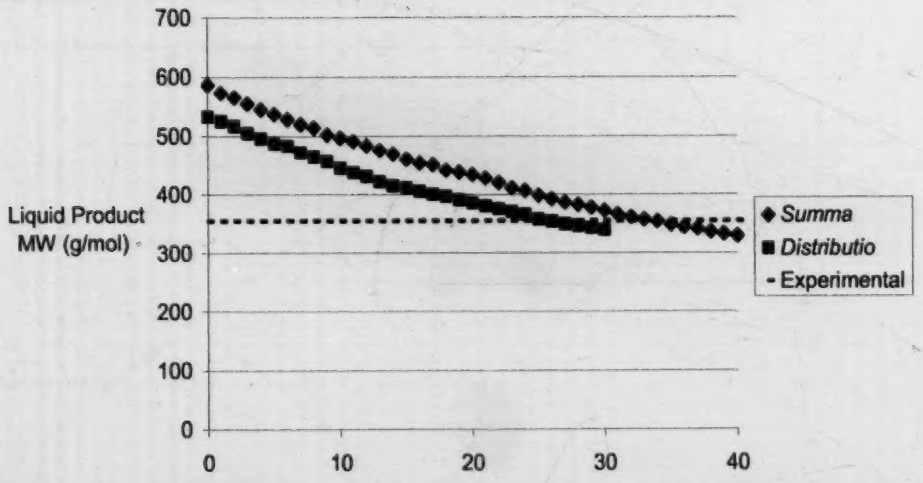


Figure 5-29. Changes in the liquid fraction average molecular weight during visbreaking simulations that used the *Summa* and *Distributio* molecular representations.

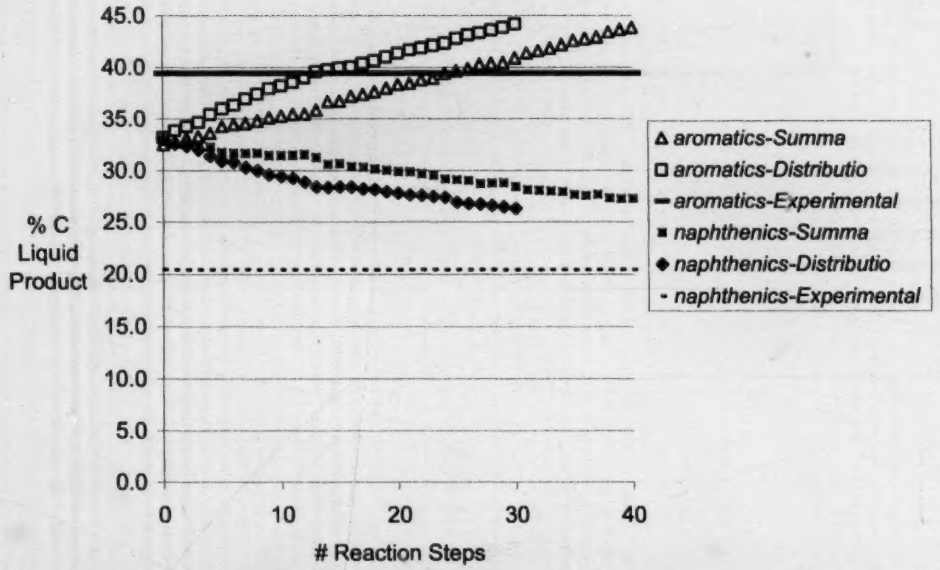


Figure 5-30. Changes in the liquid fraction aromatic and naphthenic contents during thermal cracking simulations that used the *Summa* and *Distributio* molecular representations.

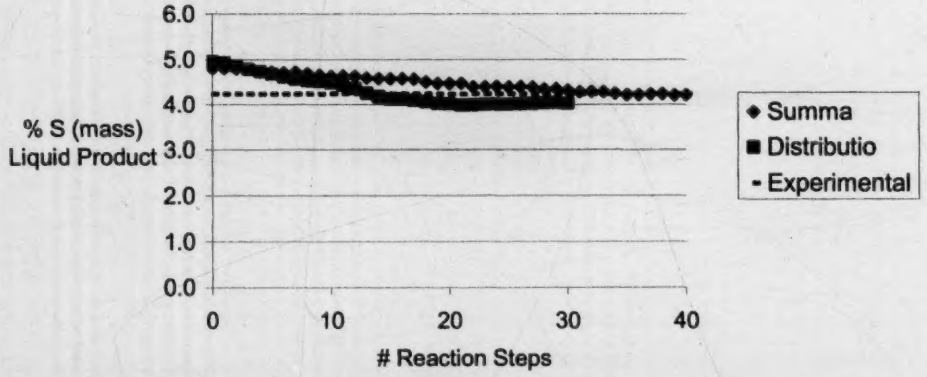


Figure 5-31. Changes in the liquid fraction sulfur composition during thermal cracking simulations that used the *Summa* and *Distributio* molecular representations.

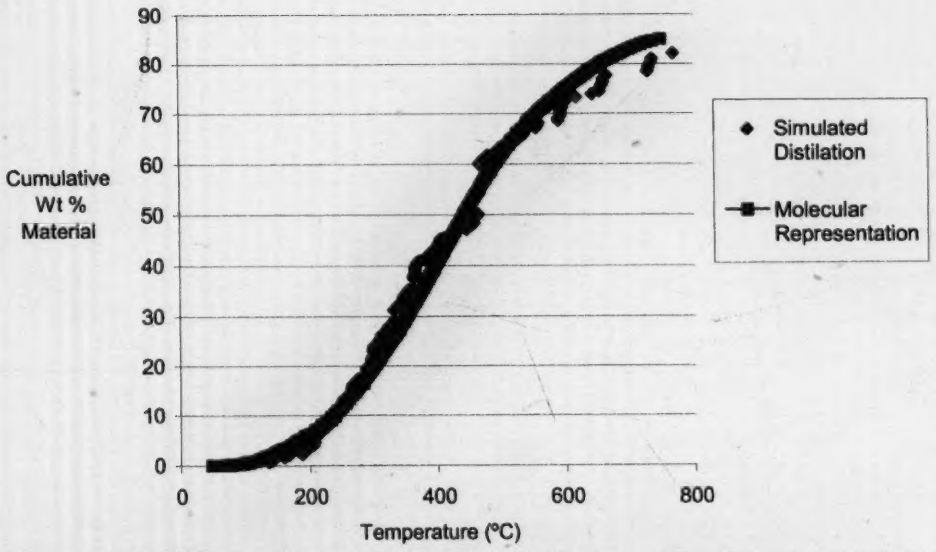


Figure 5-32. Boiling point distribution of liquid fraction molecules of a simulated *Summa* cracked product plotted versus the experimental SIMDIS data.

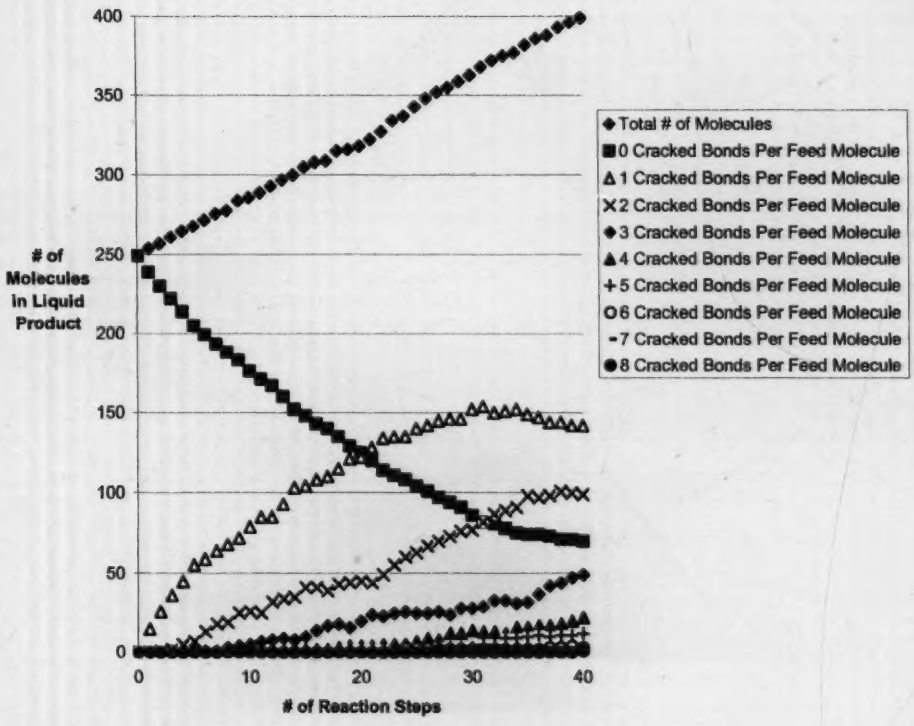


Figure 5-33. Cumulative distribution of molecules in the liquid product fraction of the *Summa* molecular representation during a thermal cracking simulation.characterized by the number of bonds cracked/feed molecule (in different reaction steps).

5.7. Discussion

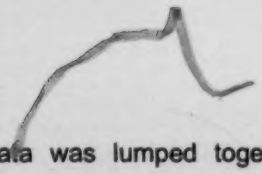
This study demonstrates the usefulness of incorporating simulated distillation data into the molecular representation optimization. The Marrero-Gani group contribution method allowed molecular representations to be created with boiling point distributions that were consistent with simulated distillation data.[6] The reactivaties and chemical properties of individual molecules varies significantly. In order to accurately model thermal cracking, a boiling point calculation method must be used since boiling point distribution data is the most readily available data for fractional characterization. Optimizing simulated distillation data creates some complications. First, it dramatically increases the computational workload. At the time (July 2007) that this was written, when simulated distillation data was optimized, it took about eight hours of optimization time to create a final molecular representation containing 15 molecules that were chosen from an initial population of 200 molecules. Conversely, when simulated distillation data was not optimized, it took about an hour of optimization time to create a molecular representation containing 15 molecules.

In Chapters 3 and 4, when simulated distillation data was not optimized, it was determined that six molecules were necessary to create a molecular representation that matched the experimental data. Conversely, when simulated distillation data was optimized, fifteen molecules were necessary to create a molecular representation that closely matched the experimental data. Thus, when more data is optimized, more molecules are required in the molecular representation. The number of molecules necessary to represent the bitumen fractions in this study are consistent with the results of Campbell and Klein.[7] In the Campbell and Klein molecular representation study, 23 molecules were required to represent a Maya residue with an average molecular weight of 944 g/mol. In order to achieve a good fit with the NMR spectroscopy, elemental, and simulated distillation data, it was necessary to use a least squares objective

function. Conversely, when simulated distillation data was not included, L1 (absolute value) objective functions offered the best performance.

The data from the thermal cracking simulations indicate that the physical behavior and properties of a petroleum feed might be better captured in molecular representations when the feed is modeled as a whole rather than in fractions. In this study the *Summa* representation both fit the experimental data better and had better performance in the cracking simulations, relative to the *Distributio* representation. In the cracking simulations, the *Summa* molecular representation performed better than the *Distributio* representation. This is an interesting result. Both the *Summa* and the *Distributio* molecular representation closely fit the simulated data. However, the *Summa* molecular representation fit the other experimental data – elemental analysis, NMR spectroscopy, and VPO – better than the *Distributio* representation. The *Summa* molecular representation was composed of 17 unique molecules while the *Distributio* representation was composed of 45 unique molecules. This study indicates that in molecular representations, the closeness of fit to the experimental data, can have a significant influence on the physical behavior of the molecular representations. In the future this could easily be tested. Molecular representations containing the same number of molecules but having varying degrees of agreement to the experimental data could be created and tested in thermal cracking simulations.

Error propagation might be responsible for the differences of the closeness of fit to the experimental data of the *Summa and Distributio* molecular representations. The molecular representation technique – including both the construction and optimization methods – is associated with a given level of error. Due to limitations, biases, and errors in the technique (that are hard to quantify), every time a molecular representation is created, a given level of error will exist between the experimental and simulated data. The *Distributio* molecular representation was composed of three independent representations: the Athabasca residue, middle distillate, and naptha. Conversely, the experimental



data was lumped together, and the *Summa* representation for the overall Athabasca feed was modeled with a single molecular representation. Thus, the error associated with the molecular representation technique would be lower in the *Summa* compared to the *Distributio* molecular representation.

Likewise, the results from the thermal cracking simulations indicate that a high quality molecular representation can be created using a minimum number of molecules. The *Summa* and *Distributio* molecular representations both fit the experimental data well, and had visbreaking simulation product yields that were consistent with the experimental yields and molecular weights. A criticism of surrogate heavy oil molecular representations in the past is that although they can model the experimental data, they will not be capable of representing other physical properties or behavior. The simulated cracking data in this study indicate that surrogate molecular representations can be created that both satisfy analytical and reactivity data. Furthermore, the thermal cracking simulations illustrate that archipelago bitumen molecules produce the full range of liquid products that are observed experimentally.

The results of the thermal cracking simulations indicate that even in mild thermal conversion processes such as visbreaking, it is necessary to break multiple bonds in each feed molecule in order to achieve a significant degree of conversion. In this study, when the restriction that prevented product molecules from reacting was lifted, the agreement between the properties of the simulated and experimental reaction products greatly improved. In the final *Summa* simulated reaction product, the majority of the product molecules were created from two or more successive cracking events.

The complexity of the chemical nature of heavy oils combined with the limitations of experimental methods, including NMR spectroscopy, make it very challenging to create accurate representations of heavy oil at the molecular level. Thermal cracking modeling represents an excellent methodology to verify

proposed chemical structures. An accurate chemical representation must satisfy both the experimental chemical (such as elemental analysis and NMR spectroscopy) and reactivity data (distillation distribution after thermal reaction). The relative errors associated with the overall chemical experimental data (due to the high NMR spectroscopy analysis error) are very high and those associated with the reactivity data are relatively low. Thus, cracking data represent a low error method of evaluating the consistency of a proposed heavy oil structure. Errors or biases in proposed structure(s) will show up as biases in the reactivity data. An iterative approach is required. One must first create a molecular representation and then react it in the thermal cracking simulation algorithm. The next step is to then compare the boiling distribution of the products to the experimentally observed boiling range. The next step is to then modify the molecular representation construction algorithm, react it again in a simulation, and so on.

Thermal cracking studies are valuable in validating the chemical structure of asphaltenes and heavy oils. Molecular representations must be capable of being consistent with all available chemical and reactivity data. The results in this study indicate that heavy oil molecules with the archipelago framework can both satisfy both the analytical and visbreaking data. The simulated cracked liquid product distribution closely matched the experimental product distribution. The simulated cracked liquid product also had an average molecular weight consistent with the experimentally determined molecular weight. The sulfur content and aromaticity of the simulated liquid product were also in close agreement with the experimentally determined values.

The thermal cracking simulations consistently produced more gas than what was observed experimentally (this is detailed by the mole balance presented in Table 5-16). In the reactivity simulations, relative to the experimental data, approximately three times as much gas was produced. In Table 5-16, it is also apparent that relative to the experimental data, about 1.75

times the total number of cracking reactions occurred in the simulations. However, the composition of molecules produced in the simulated liquid fraction was comparable to the experimentally predicted value. Thus, in the cracking simulations gas formation is too favorable – or too many cracks produce gas molecules. Possibilities for this include:

1. Too many aromatic side chains in the molecular representations.
2. Aromatic sidechains that were too long.
3. Aromatic groups that were connected by C-C bridges that were too long in the molecular representations.
4. Certain relative cracking propensities that were too high.

Possibilities 1-4 would all result in a higher percentage of cracking events that produce gas. During visbreaking, side chains will crack off the aromatic groups. Likewise, C-C chains between aromatic groups will crack to yield aromatic groups with side chains. Thus, as the aromatic side chains in a molecular representation are increased in length and/or abundance, the amount of gas production in visbreaking will increase. Thus, in the future it would also be recommended to include constraints that can limit the length and abundance of aromatic side chains, and constraints that can restrict the length of the C-C chains between aromatic groups. Likewise, if the relative cracking propensities associated with aromatic sidechains were decreased, gas formation during simulation will also decrease.

In this study, the relative yield of reaction products published in the literature were used to estimate relative cracking propensities (RCP). The RCPs were then used in stochastic reactivity simulations. Conversely, Neurock and Klein used first rate reaction constants to estimate the relative probabilities of reaction[8]. In the future, it will be recommended to estimate the RCPs using kinetic rate constants instead of product distributions yields. This will add a significant degree of complexity to the model, however it will improve the product

prediction capability of the model, and it will allow a kinetic model to be developed. When values are obtained for the rate constants for each reaction, the time associated with each reactive event can also be calculated.

In heavy oil thermal cracking, methane is formed via CH_3 radicals that are created from β scission reactions of intermediate radicals – primarily from radicals that exist in the gas phase. Thermodynamically, CH_3 radicals can also be produced from initiation reactions. However, for reactants larger than C_4 this is a thermodynamically rare event. In this study, it was assumed that gas phase molecules did not react. Also, the relative cracking propensity approach essentially lumped all of the reaction steps together. Thus, the mechanism of methane production in the simulations in this study is very different from experimental/real life methane production. On a production per feed basis, the methane produced in the simulations was consistent with the experimentally determined value. During the simulations, when a terminal (CH_2-CH_3) bond had the potential to produce methane as a product, a relative cracking propensity was chosen relative to a phenyl, pyrene, branch CH, naphthenic ring, or thioether S group. In the future it would be recommended to calculate relative cracking propensities of terminal (CH_2-CH_3) bonds using data based upon the experimental production of methane.

Although the aromaticies of the simulated cracked liquid products were consistent with the experimentally determined value, the naphthenic contents were not in agreement. However, it is difficult to determine if the disagreements between the naphthenic contents of the simulated liquid products and the experimentally determined values were due to errors in the modeling of the naphthenic structures in the molecular representations, errors in the modeling of the transformation of naphthenic structures into aromatic structures, or a combination of both errors in the naphthenic structural and cracking modeling. In the future, insights will be obtained into this conundrum if visbreaking simulations are carried out on molecular representations that contain alternative naphthenic

structures. Likewise, the distribution between aromatic associated naphthenics and aliphatic naphthenics should be varied in the molecular representations that are used in future visbreaking simulations. In this study, a constant fractional transformation of 50% at each reaction step of specific naphthenic structures into aromatic structures was used. Alternative approaches to modeling the naphthenic to aromatic transformations should be pursued.

5.8. References

1. Kissin, Y.V. *Ind. Eng. Chem. Res.* **1987**, *26 (8)*, 1633-1638.
2. Abikhers, V.; Fixari, B.; Leperchex, P. *Fuel.* **1986**, *65 (3)*, 442-446.
3. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1987**, *26(2)*, 374-376.
4. Freund, H.; Matturro, M.G.; Olmstead, W.N.; Reynolds, R.P.; Upton, T. H. *Energy & Fuels.* **1991**, *5(6)*, 840-846.
5. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1988**, *27*, 1348-1356.
6. Marrero, J.; Gani, R.; *Fluid Phase Equilibria*, **2001**, *183-184*, 183-208.
7. Campbell, D M.; Klein, M.T. *Applied Cat. A: General* **1997**, *160*, 41-54.
8. Neurock, M.; Libanati, C.; Klein, M.T. Modelling asphaltene reaction pathways: intrinsic chemistry. *AIChE Symposium Series.* **1989**, *273(85)*, 7-14.

Chapter 6. Discussion and Recommendations

The molecular representations and thermal cracking simulations presented in this thesis provide support for the archipelago structure of heavy oils and asphaltenes. The molecular representations for asphaltenes and a various residue fractions were all consistent with a wide range of experimental data: NMR spectroscopy, elemental analysis, vapor pressure osmometry, and simulated distillation. In the research presented in this thesis, bitumen molecular representations containing archipelago structures produced liquid product fractions with mass distributions that closely matched experimental data. Pericondensed structures can not explain the full range of products that occur during thermal cracking of bitumens. However, further simulations, experiments and scientific developments are necessary to finally resolve the structural chemistry of asphaltenes and heavy oils.

The work and data documented in this thesis indicate that molecular representations that contain a minimum number of molecules can be created for heavy oil fractions. The molecular representations of the Athabasca bitumen fractions were consistent with a plethora of analytical data: NMR spectroscopy, elemental analysis, vapor pressure osmometry, and simulated distillation. Critics of molecular representations point out that even the NMR spectroscopy analysis of simple molecules is very challenging and error prone. They reason that it is therefore not applicable to use such a technique on nonhomogenous mixtures of compounds such as heavy oils. However, the approach that was used in this study assured that the concentrations of molecules were adjusted so that the calculated properties of molecules in the molecular representations were consistent with all of the data being modeled – not just NMR spectroscopy. Thus, errors, biases, or shortcomings associated with a specific analytical technique will be minimized due to the mathematical constraints imposed by including data from other analytical methods. The method presented in this

thesis is versatile; it is very straightforward to take into account new data and include it in a molecular representation. New data can either be incorporated at the construction stage, at the optimization stage, or in both.

True to their name, the Monte Carlo methods developed in this thesis are random methods. Thus, there is considerable variation amongst the molecules produced from the construction algorithm. The work in this thesis indicates that in order to produce a molecular representation that is consistent with the analytical data, the initial starting population of molecules must be sufficiently large. This is expected since the construction algorithm will produce both "good" (molecules that are have properties relatively close to the experimental properties) and "bad" molecules (molecules that have properties that significantly deviate from experimental properties). The greater the number of molecules in the initial population of molecules, the more "good" molecules will be present. A certain number of "good" molecules are definitely required to produce a high quality molecular representation that is consistent with experimental data. Likewise, it certainly appears that when more data is modeled/fitted, more molecules are required in molecular representation. The molecular representations that modeled the simulation data required more than twice as many molecules than those that did not model the simulated distillation data.

An important question arises from this work – is it worth the time and effort to create molecular representations? Do molecular representations provide value that is not present in the raw analytical data? From a scientific perspective, molecular representation is a methodology that allows researchers to gain insight into the molecular structure of bitumens and asphaltenes. With knowledge of the true chemical structure of bitumens and asphaltene, scientists and engineers will be able to design and implement more efficient and effective technologies to recover and process these fuels. Likewise, as presented in Chapter 5, molecular representations allow the development of molecular level reactivity simulations. Molecular level reactivity simulations have the potential to accurately simulate the

upgrading and refining of heavy oils. Likewise, group contribution theory allows various physical properties to be calculated from molecular representations – that can not easily be obtained from experiments or raw analytical data. Group contribution theory enables the calculation of critical properties such as T_C, P_C, and V_C of individual molecules. This opens up the door to equation of state modeling that allows the estimation of density and viscosity values. Thus, molecular representations have implicit value in both the upstream and downstream segments of the oil industry.

The thermal cracking results presented in Chapter 5 illustrate both the advantages and disadvantages of the addition of complexity to molecular representations. Including simulation distillation data produced molecular representations that modeled bitumen the molecular weight distribution. The molecular representations that were created with simulated distillation data accurately predicted the distribution of liquid products in the visbreaking simulations. However, the molecular representation created from representations from the separate distillation fractions, did not perform as well in the visbreaking simulations compared to the molecular representation created from the overall combined data. This indicates that when too much complexity is included in molecular representations, an unwanted degree of error is potentially introduced into the representations. Thus, a tranquil balance must be obtained. Molecular representations must truly represent the structural chemistry of a petroleum, however, if too much information is included in a representation, an unacceptable degree of error might be introduced.

Where does one go from here? Model compounds will have a significant role in the future development of molecular representations. It is necessary to quantify the error associated with the NMR spectroscopy analysis of heavy oil fractions. The NMR spectroscopy analysis of both compounds similar to those in the reactive studies featured in Chapter 2 and of those of the larger asphaltene like compounds will allow the measurement and reproducibility error of NMR

spectroscopy measurement to be quantified. Likewise, in the future when synthetic bitumen model compounds are available, they will be used as controls in both molecular representations and thermal cracking modeling. Molecular representations of synthetic bitumen model compounds will allow molecular representation methodology to be tested. In principle, molecular representations should be very accurately created for mixtures of known structure and composition. Likewise, thermal cracking studies with synthetic bitumen model compounds will allow more accurate molecular representation cracking simulations to be developed.

The thermal cracking simulations featured in Chapter 5 need to be refined into kinetic simulations. Kinetic rate constants from model compound kinetic studies will be required to accomplish this. A kinetic approach will allow detailed product prediction during thermal cracking. A kinetic approach has the potential of more accurately predicting the rate of gas formation. Likewise, a low degree of conversion was modeled in this study. In the future, it will be interesting to see what happens to the molecule requirements when higher conversion degrees are modeled with and without kinetics.

Molecular representations represent a technique that can yield physical and chemical information about petroleums while decreasing the requirement for experimentation. The advantage of molecular representation is that it represents petroleum at the most fundamental level: the molecule. The fundamental nature of molecular representations allows physical properties to be calculated, while serving as the basic framework for process models that require input at the molecular level.

This thesis illustrates the value of heavy oil molecular representations. Heavy oil molecular representations represent a new technique that is evolving rapidly. The usefulness of molecular representations in structural chemistry, cracking, and physical property calculation have been illustrated in this thesis.

The cracking and structural uses of molecular representations are currently being used within industry to gain insights into the molecular nature of petroleum. New theories, correlations, and models are currently being developed that will extract even more information from molecular representations. The future is very bright in the molecular representation world.

Chapter 7. Conclusions

Molecular representations that contained fewer than 20 molecules modeled both the analytical characterization and cracking data of Athabasca residue fractions. The molecular representations presented in this thesis had very good agreement with NMR spectroscopy, elemental analysis, vapor pressure osmometry, and simulated distillation data. Thus, thousands of molecules are not necessary to model the structural chemistry of a heavy oil feed. However, as more data is modeled, more molecules are required (when simulated distillation data was incorporated, more molecules were necessary). A significantly large starting population of at least 50 molecules must be created in order to create high quality molecular representation that contains relatively few molecules. At least five molecules are required to produce a molecular representation that is consistent with experimental data.

The results of the thermal cracking simulations indicate that in mild thermal conversion processes such as visbreaking, it is necessary to break multiple bonds in each feed molecule in order to achieve a significant degree of conversion. The data from the thermal cracking simulations indicate that the physical behavior and properties of a petroleum feed might be better captured in molecular representations, when a feed is modeled as a whole rather than in fractions.

The thermal cracking simulations indicate that high quality molecular representation can be created using a minimum number of molecules. In the thermal cracking simulations, the bitumen molecular representations produced liquid fractions with product compositions similar to those that were observed experimentally. The liquid mass and boiling point distributions of the simulated cracked liquid products were consistent with the experimental data. Likewise, the average predicted liquid fraction molecular weights were consistent with the experimentally determined value.

Thus, the results in this thesis demonstrate that heavy oil molecules with the archipelago framework can both satisfy analytical characterization and visbreaking data. However, this does not mean that the structures proposed in this thesis are in fact the correct structures. More experiments and more modeling will be required to verify the structures proposed in this thesis.

Appendix A. Model Compound Thermal Cracking Studies

A.1. Branched Model Compounds

Kissin studied the free radical cracking at 250 °C of three branched alkanes: pristane, phytane, and squalene (Figure A-1).[1] A capillary gas chromatogram was used to identify all of the reaction products at 250 °C. Kissin was able to explain the cracking process using only three reaction steps: hydrogen abstraction reactions that form parent radicals, beta-scission reactions of the parent radicals, and chain transfer reactions that involve both substrate molecules and all formed radicals. By analyzing the identified reaction products, Kissin was able to calculate the probabilities of both radical formation and bond cleavage of atoms in the branched alkane feed. The probabilities of breakage of the CH_2-$CH(CH_3)$,CH_2-CH_2 and CH_2-CH_2 had ratios of: 1.5:1:0.1.

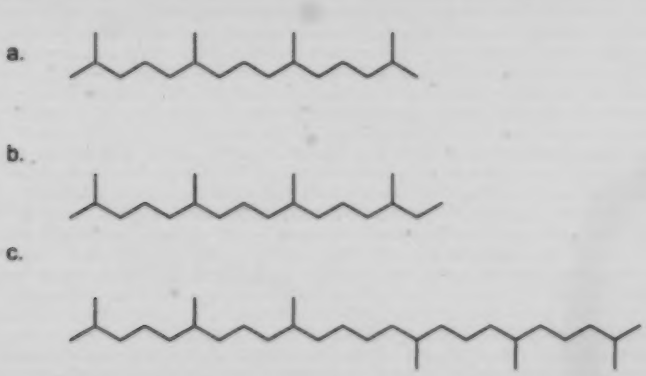


Figure A-1. The structures of pristane (a.), phytane (b.), and squalane (c.).

A.2. Sulphur Containing Model Compounds

Abikers *et al.* studied the thermal reactivities of didoceylsulphide and dodedecylsulphone (Figure A-2).[2] Didoceyl sulfide contains two, twelve carbon linear alkane chains connected together by a sulfur atom. Didoceyl sulfide is a twelve carbon alkane chain connected to a sulphur atom radical. These two compounds were pyrolyzed at both 500 °C and 530 °C, and the products were identified using gas chromatography. The pyrolysis reactions were carried out in the presence of various solvents having different H-transfer power. Abikers *et al.* found that as the hydrogen transfer ability of the medium increased, the degree of conversion decreased. At 500 °C, a 48% degree of conversion was measured for the pyrolysis reaction in benzene. Conversely, the measured degrees of conversion in tetralin and tetralin - DHA were 40% and 35% respectively. Only traces of n-alkane products were measured. Significantly more C_{12} olefin products were measured than the combined C_6-C_{11} products: a ratio of 3.88:1 was measured for the pyrolysis in tetralin at 500 °C. This corresponds to a ratio of about 19.4:1 that compares the C_{12} products to each C_6 to C_{11} product. In this study, the concentrations of C_6 and C_7 products appear to be unusually higher than expected. These high concentrations might be due to an experimental flaw as they can not be explained using conventional reaction mechanisms. The distribution of n-olefin products from didoceyl sulfide pyrolysis at 500 °C is presented in Table A-1.

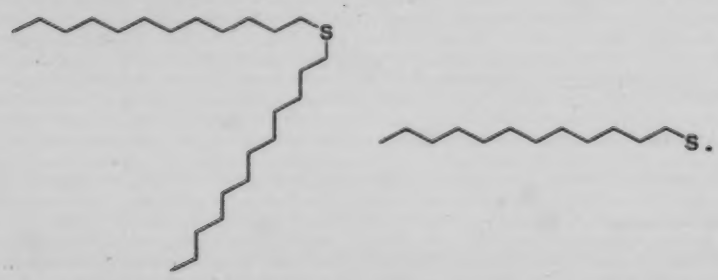


Figure A-2. The structures of didodecyl sulfide (a.) and 1-dodecanethiol (b.).

Table A-1. Thermal cracking of n-didodecyl sulfide.[2]

Temperature (℃)	Diluents	Conv (%)	C_1-C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
500	Benzene	48	10	8	9.9	4.9	5.0	4.3	4.0	23.5
500	Tetralin	40	8	4.5	4.2	1.8	2.1	1.8	1.9	48.5
500	Tetralin-DHA	35	3	4.3	2.5	2.0	1.8	1.9	2.0	47.2
530	Benzene	90	20	12.5	5.1	5.1	6.0	5.2	5.0	25.3
530	Tetralin	82	14	10.1	5.8	4.8	6.3	5.0	4.7	36.2
530	Tetralin-DHA	80	6	9.6	6.0	3.9	4.8	3.9	3.9	53.5

A.3. Phenyl Model Compounds

Savage and Klein studied the pyrolysis of phenyldodecane at 400 °C in tetralin (Figure A-3).[3] The liquid products of the reactions were analyzed using gas chromatography. The major product of the pyrolysis reactions was toluene. The molar yields of the pyrolysis products of phenyldodecane are presented in Table 3-2. Deuterium labeling proved that the pyrolysis of phenyldodecane involved a free radical mechanism and that the pyrolysis mechanism was entirely free radical.

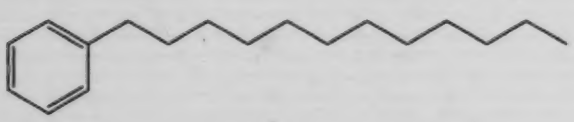


Figure A-3. The structure of phenyldodecane.

Table A-2. Molar yields of products from phenyldodecane thermal cracking at 400 °C.[3]

	molar yield (%) at reaction times (min)				
product	30	60	90	120	240
hexene	0.5	0.75	0.94	1.42	1.71
hexane	0.4	0.61	1.03	1.39	3.37
benzene	0.1	0.18	0.26	0.44	0.69
heptene	0.35	0.55	0.72	1.04	1.28
heptane	0.29	0.52	0.92	1.19	2.69
toluene	4.48	7.88	12.47	15.95	28.34
octene	0.35	0.55	0.69	0.97	0.92
octane	0.12	0.26	0.5	0.64	1.61
ethylbenzene	0.36	1.04	2.41	2.43	7.55
nonene	0.2	0.29	0.35	0.47	0.43
styrene	0.84	0.93	0.73	1.04	0.36
nonane	0.22	0.47	0.79	1	1.79
phenylpropene	0	0.08	0.08	0.12	0.09
propylbenzene	0.14	0.25	0.48	0.62	1.49
decene	0.28	0.37	0.42	0.54	0.44
decane	2.18	3.26	4.85	5.5	7.76
butylbenzene	0.19	0.4	0.68	0.86	1.66
undecene	3.28	4.03	3.86	4.54	2.27
undecane	0.18	0.24	0.41	0.56	1.22
phenylpentene	0.09	0.12	0.12	0.14	0.09
phenylpentane	0.2	0.34	0.61	0.77	1.42
dodecane	0	0.04	0.09	0.12	0.3
phenylhexene	0.11	0.14	0.16	0.2	0.22
phenylhexane	0.15	0.24	0.42	0.51	0.9
phenylheptene	0.12	0.13	0.13	0.15	0.12
phenylheptane	0.13	0.21	0.38	0.44	0.72
phenyloctene	0.09	0.1	0.1	0.11	0.07
phenyloctane	0.16	0.25	0.41	0.47	0.71
phenylnonene	0.1	0.11	0.11	0.12	0.09
phenylnonane	0.18	0.28	0.45	0.51	0.73
phenyldecene	0.19	0.28	0.35	0.47	0.28
phenyldecane	0.09	0.17	0.28	0.38	0.41
phenylundecene	0.03	0.09	0.08	0.09	0.12
phenylundecane	0.15	0.21	0.23	0.26	0.33
phenyldodecane	93.26	75.48	54.25	50.67	22.22

Savage and Klein also studied the pyrolysis of n-pentadecylbenzene (PDB) at temperatures between 375 to 450 °C (Figure A-4).[4] The major products of the pyrolysis reactions were toluene, 1-tetradecene, ethylbenzene, and n-tridecane. After 60 minutes at 450 °C the relative yields were: toluene (64.5%), ethylbenzene (27.2%), styrene (0.39%) and n-tridecane (1.32%). Pyrolysis in deuteriated tetralin revealed that n-pentadecylbenzene was broken down entirely by a free radical pathway. No benzene formation or C15 products were reported by Savage and Klein.

The work of Savage and Klein was in agreement with the study by Mushrush and Hazlett.[5] Mushrush and Hazlett studied the pyrolysis of both n-pentadecylbenzene and 2-n-pentadecylpyridine at 450 °C (Figure A-4). There were no C_{15} products measured. At 180 minutes, the yields of toluene and ethylbenzne - the two major products - produced from n-pentadecylbenzene were 14.9% and 0.2% respectively. In the case of 2-n-pentadecylpyridine, the yields of toluene and ethylbenzne were 14.4% and 0.2% respectively. It was reported that free radical formation is greatly enhanced by pyridine and benzene rings. It took 120 minutes to decompose the equivalent amount of n-hexadecane that was decomposed by 2-n-pentadecylpyridine in 30 minutes.

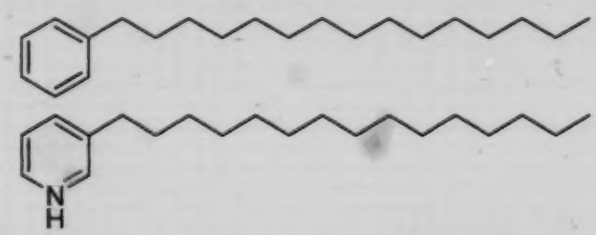


Figure A-4. The structures of: a. pentadecylbenzene and b. 2-n-pentadecylpyridine.

A.4. Pyrene Model Compounds

Freund *et al.* studied the thermal cracking of 1,20-di(1-pyrenyl)eicosane at both 427 °C and 512 °C (Figure A-5).[6] A major assumption of previous studies was that the number of rings does not influence the pyrolysis reactions that occur. The conclusion of this study was that pyrene sidechain pyrolysis is highly unusual; the pyrene structure has an actual role in the pyrolysis. The observed bond cleavage of the aryl-alkyl C-C bond was significantly greater than what is observed for one and two ring aromatic systems. Analysis of the products led Freund *et al.* to conclude that the pyrolysis reactions involve the formation of an internal olefin conjugated to pyrene. This olefin intermediate provides hydrogen to cleave bonds. The direct cleavage at the pyrene ring was preceded by a phenalenyl-like radical. The alkyl chain length distribution of the pyrolysed products of 1,20-di(1-pyrenyl)eicosane is presented in Table A-3.

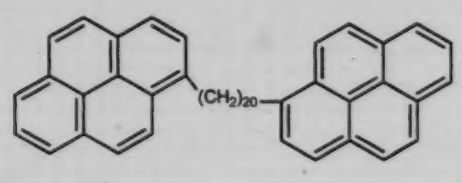


Figure A-5. The structure of 1,20-di(1-pyrenyl)eicosane.[6]

Table A-3. Product selectivities for the thermal cracking of 1,20-di(1-pyrenyl)eicosane at 427 °C.[6]

Alky Chain Length on Pyrene Core	Relative Amount of Product
0	12
1	11
2	4.9
3	2.4
4	5
5	1
6	2.6
7	1.5
8	1
9	2.6
10	1
11	1
12	1
13	1.5
14	2
15	1.5
16	2.5
17	6
18	8
19	15
20	17.5

Savage *et al.* pyrolyzed 1-dodecylpyrene (DDP) at temperatures from 375 °C to 450 °C (Figure A-6).[7] At low DDP concentrations, DDP pyrolysis was determined to occur via a pathway that is analogous to the pathway responsible for the pyrolysis of n-alkylbenzene. However, at higher concentrations, an autocatalytic pathway that cleaves aryl-alkyl C-C bonds is dominant. The molar pyrene yields that were measured after complete pyrolysis were 0.08, 0.18, and 0.45 for initial DDP concentrations of 0.005, 0.03 and 2.3 M. Thus, the work of Savage and Klein and Freund *et al.*, indicates that alkyl-pyrene pyrolysis involves cleavage of aryl-alkyl C-C bonds.

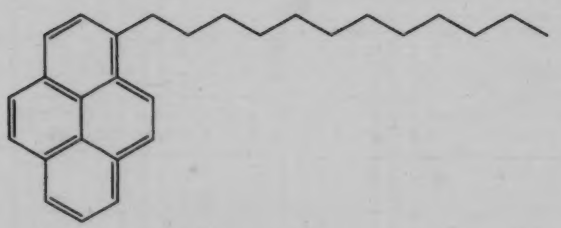


Figure A-6. The structure of 1-dodecylpyrene.[7]

Smith and Savage studied the pyrolysis of both 2-(3-phenylpropyl) naphthalene (PPN) and 1,3-bis(1-pyrene)propane (BPP) at temperatures ranging between 365 °C and 450 °C (Figure A-7).[8] The major products of PPN pyrolysis were: 2-methylnaphthalene, styrene, toluene, and 2-vinylnaphthalene. Benzene and naphthalene production were not recorded. Conversely, pyrolysis of BPP led to the production of vinylpyrene, methylpyrene, and pyrene-as a product with a 36% yield produced by secondary reactions. Thus, in this study aryl-alkyl C-C bond cleavage was observed for pyrene, but not for benzene or naphthalene. However, for both PPN and BPP, Smith and Savage concluded that the major reaction pathway involved a Rice-Herzfeld free radical chain mechanism.

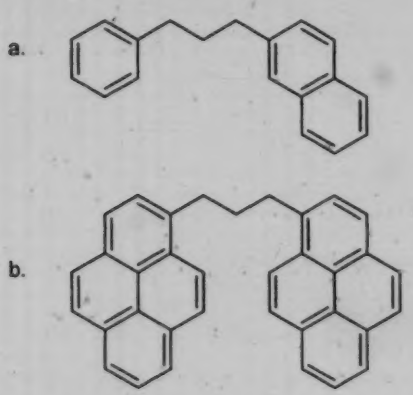
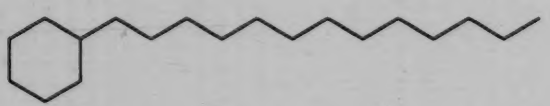


Figure A-7. The structures of: a. 2-(3-phenylpropyl) naphthalene. b. 1,3-bis(1-pyrene) propane.

A.5. Naphthenic Model Compounds

Savage and Klein also studied the pyrolysis of tridecylcyclohexane and 2-ethyltetralin (Figure A-8)[9]. Tridecylcyclohexane and 2-ethyltetralin were independently pyrolysed at 400 °C, 425 °C, and 450 °C in tubing bomb reactors. The reaction products of the pyrolysis reactions were analyzed by gas chromatography and GC-MS. The major products of the reaction were cyclohexane, dodecane, methylene cyclohexane, and tridecene. Toluene formation was only observed at 450 °C. A complete list of the reaction products from the pyrolysis of tridecylcyclohexane is presented in Table A-4. The major products of 2-ethyltetralin pyrolysis were naphthalene, 2-ethyldialin, 2-ethyl naphthalene, dialin, and tetralin. A complete list of the reaction products from the pyrolysis of 2-ethyltetralin is presented in Table A-5.

a.



b.

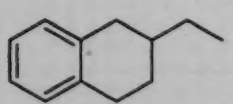


Figure A-8. The Structures of: a. Tridecylcyclohexane and b. 2-Ethyltetralin.

Table A-4. Molar yields of products from tridecylcyclohexane thermal cracking at 400 ℃.[9]

Time (min)	30	90	180
cyclohexane	0.59	1.58	3.14
cyclohexene	0.27	0.73	1.4
heptene	0.09	0.43	0.68
heptane	0.1	0.54	1.26
methylcyclohexane	0.22	0.65	1.49
methylenecyclohexane	0.63	1.01	1.96
toluene			
methylcyclohexenes	0.1	0.81	1.1
octene	0.06	0.17	0.4
octane	0.13	0.43	1.01
cyclohexylethene	0.17	0.39	0.61
ethylcyclohexane	0.18	0.54	1.17
nonene	0.06	0.25	0.44
nonane	0.07	0.39	0.94
cyclohexylpropene	0.15	0.36	0.51
propylcyclohexane	0.14	0.42	0.98
decene	0.05	0.18	0.32
decane	0.07	0.4	0.93
cyclohexylbutene	0.12	0.27	0.33
butylcyclohexane	0.11	0.35	0.75
undecene	0.05	0.17	0.27
undecane	0.07	0.39	0.87
cyclohexylpentene	0.1	0.24	0.28
pentylcyclohexane	0.13	0.38	0.82
dodecene	0.06	0.2	0.27
dodecane	0.55	1.46	2.91
cyclohexylhexene	0.06	0.23	0.26
hexylcyclohexane	0.13	0.37	0.76
tridecene	0.34	0.66	0.75
tridecane	0.19	0.55	1.12
cyclohexylheptene	0.12	0.22	0.24
heptylcyclohexane	0.17	0.42	0.83
cyclohexyloctene	0.05	0.17	0.21
octylcyclohexane	0.11	0.33	0.67
cyclohexylnonene	0.05	0.15	0.14
nonylcyclohexane	0.1	0.3	0.58
cyclohexyldecene	0.05	0.13	0.13
decylcyclohexane	0.26	0.46	0.68
tridecylcyclohexane	87.4	88.8	61.1
alkanes	1.18	4.15	9.05
olefins	0.75	2.06	3.14
alkylcycolohexanes	2.15	5.8	11.9
cyclohexylalkenes	1.77	3.9	6.07

Table A-5. Molar yields of products from 2-Ethyltetralin thermal cracking at 400 ℃.[9]

time (min)	10	20	30	45	60	90	120
benzene	0.03	0	0.03	0.02	0.03	0.01	0.02
toluene	0.19	0.1	0.23	0.2	0.25	0.18	0.28
methylindan	0	0	0.03	0.05	0.07	0.08	0.15
tetralin	0.12	0.17	0.31	0.46	0.53	0.63	0.97
dialin	1.34	1.09	1.39	1.37	1.53	1.2	1.6
naphthalene	1.01	1.13	1.67	2.21	2.52	2.75	4.08
2-ethyldialins	1.84	1.19	1.03	0.71	0.59	0.32	0.27

A.6. References

1. Kissin, Y.V. *Ind. Eng. Chem. Res.* **1987**, *26 (8)*, 1635-1638.
2. Abikhers, V.; Fixari, B.; Leperchex, P. *Fuel.* **1986**, *65 (3)*, 442-446.
3. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1987**, *26(2)*, 374-376.
4. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.***1987**, *26(3)*, 488-494.
5. Mushrush, G.W.; Hazlett, R.N. *Ind. Eng. Chem. Fundam.* **1984**, *23(3)*, 288-294.
6. Freund, H.; Matturro, M.G.; Olmstead, W.N.; Reynolds, R.P.; Upton, T. H. *Energy & Fuels.* **1991**, *5(6)*, 840-846.
7. Savage, P.E.; Jacobs, G.E.; Javanmardian, M. Ind. Eng. Chem. Res. **1989**, *28*, 645-654.
8. Smith, C.M.; Savage, P.E. *Energy & Fuels.* **1991**, *5(1)*, 146-155.
9. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1988**, *27*, 1348-1356.

Appendix B. The First Molecular Representation Creation Algorithm

This appendix provides a detailed description of the first molecular representation creation algorithm presented in Chapter 3.

To reduce the computational workload, two different types of chain methylene groups were used: chain methylene and extended chain methylene. The chain methylene groups are individual CH_2 carbons. The extended chain methylene groups are clusters of three CH_2 carbons that are connected together.

The following is a summary of the Monte Carlo construction method used in this study:

1. The concentrations of heteroatom and metal containing groups, including indole, dibenzothiophene, dibenzofuran, and porphyrin, are calculated using the elemental concentrations. It is assumed that 60% of the sulfur is associated with the aromatic rings and the balance is associated with thioether bridges. All the nitrogen is assigned to aromatic rings. It is assumed that all the oxygen is present in aromatic rings and all the vanadium is present within porphyrin rings. The aromatic ring groups used in this algorithm are presented in Figure 1-18. The concentrations of benzene, naphthalene, phenanthrene, phenylnaphthalene, biphenylnaphthalene and pyrene are optimized using nonlinear optimization. The concentration of each of these aromatic groups is optimized so that the total Q_1, Q_2, C_1 and C_2 concentrations calculated from all of the aromatic groups are consistent with the analytical values obtained from NMR spectroscopy. Half of the naphthenic carbons are assumed to be associated with aromatic rings and half were associated with aliphatic carbons. It is assumed that one half of terminal α-CH_3 carbon is associated with aromatics, one quarter associated

with chain methyne carbons, and one quarter associated with naphthenic rings.

2. Aromatic groups (Figure 1-18) are randomly sampled. Each type of aromatic group is allowed to have a maximum number of $2n + 1$ attachments, where n represents the number of rings in the aromatic group. Each sampled group is documented in a structural matrix.
3. For each aromatic group that was sampled in step 2. α-CH_3 carbon, β-terminal chains and α-methylene groups are randomly sampled and recorded in the structural matrix. At this point, α-methylene groups are not directly connected to the aromatic groups. They represent potential connection points to other groups. In step 11, "other carbons" are attached to aromatic groups, to represent these α-methylene groups. In step 11, the α-methylene groups will become documented in both the connection and structural matrices. During random sampling, the algorithm assures that the total number of groups sampled is less than or equal to the maximum number allowed ($2n + 1$) for each type of aromatic group.
4. For each aromatic group, naphthenic rings are sampled and documented in the structural matrix. During the naphthenic sampling, the algorithm assures that each aromatic group contains less than or equal to ($2n + 1$) total attachments.
5. The valency of each aromatic cluster is defined as:

$$v_{AR,i} = \sum \alpha - methylene_{unconnected} - \sum \alpha - methylene_{connected}$$

Where:

$\alpha - methylene_{unconnected}$ represents aromatic associated α methylene carbon that is not attached to other functional groups.

$\alpha - methylene_{connected}$ represents aromatic associated α methylene carbon that is connected to other functional groups.

The total valency of all the aromatic groups is defined as:

$$TOTALv_{AR} = \sum_{i=1}^{A} v_{AR,i}$$

Where *A* represents the total number of aromatic groups.

6. The minimum total valency required to connect all of the groups together is:

$$v_{min} = 2 + 2(A - 2)$$

If $TOTALv_{AR}$ is less than v_{min}, aromatic groups are randomly sampled, and α terminal CH_3 carbons and β-terminal chains are randomly removed to increase $TOTALv_{AR}$. This is continued until $v_{min} = TOTALv_{AR}$.

7. Chain methyne, α-CH_3 chain methyne, chain methylene and extended chain methylene groups are sampled and recorded in the structural matrix.
8. Thioether and aliphatic naphthenic rings are sampled and recorded in the structural matrix.
9. A connection matrix is created that has rows corresponding to: aromatic groups, chain methyne, α-CH_3 chain methyne-, chain methylene and extended chain methylene groups. A connection algorithm randomly connects all of the groups together subject to a few constraints:

    -Groups must only connect to other groups. They are not allowed to connect to themselves.

    -Connections are only allowed between a group that has already been connected to the developing molecule, and a group that does not have any connections. This prevents paraffinic carbon from organizing into naphthenic secondary structure.

    -Aromatic groups are only allowed to have a single biphenyl connection to another aromatic group.

-Chain methylenes are only allowed to connect to other chain methylenes and to extended chain methylene groups.

-Thiophenes are not allowed to connect to other thiophenes.

-Group valencies are not allowed to proceed below zero. Each group has a predetermined number of attachments that it is allowed to connect to.

The valencies are defined as:

$$v_{CM} = 3 - \sum \text{attached aromatic groups} - \sum \text{attached extended chain methylenes} - \sum \text{attached chainmethynes} - \sum \text{attached } \alpha - CH_3 \text{ chainmethynes} - \sum \text{attached thioethers}$$

$$v_{CM\alpha} = 2 - \sum \text{attached aromatic groupss} - \sum \text{attached extended chain methylenes} - \sum \text{attached chainmethynes} - \sum \text{attached } \alpha - CH_3 \text{ chainmethynes} - \sum \text{attached thioethers}$$

$$v_{CMethylene} = 2 - \sum \text{attached chain methylenes} - \sum \text{attached extended chain methylenes}$$

$$v_{CMethylene_{ext}} = 2 - \sum \text{attached aromatic groups} - \sum \text{attached chain methylenes} - \sum \text{attached extended chain methylenes} - \sum \text{attached chainmethynes} - \sum \text{attached } \alpha \text{ ter min al chainmethynes} - \sum \text{attached thioethers}$$

$$v_{thio} = 2 - \sum attached\ aromatic\ groups - \sum attached\ extended\ chain\ methylenes - \sum attached\ chainmethynes - \sum attached\ \alpha - CH_3\ chainmethynes$$

$$v_{aro} = \sum \alpha\ methylenes_{unconnected} - \sum \alpha\ methylenes_{connected} - \sum attached\ extended\ chain\ methylenes - \sum attached\ chainmethynes - \sum attached\ \alpha - CH_3\ chainmethynes - \sum attached\ thioethers$$

For v_i:

attached aromatic groups represent aromatic groups that are attached to group "i".

attached chain methylenes represent chain methylene groups that are attached to group "i".

attached extended chain methylenes represent extended chain methylene groups that are attached to group "i".

attached chainmethynes represent chain methynes groups that are attached to group "i".

attached chainmethyne α terminals represent α terminal CH3 chain methyne groups that are attached to group "i".

attached thioethers represent thioether groups that are attached to group "i"

v_{CM} represents the valency of a chain methyne.

$v_{CM\alpha}$ represents the valency of a α-CH3 chain methyne .

$v_{CMethylene}$ represents the valency of a chain methylene.

$v_{CMethylene_{ext}}$ represents the valency of an extended chain methylene.

v_{aro} represents the valency of an aromatic group.

v_{thio} represents the valency of a thioether.

At each step, the connection algorithm randomly samples two groups, A and B. When A and B do not conflict with any of the constraints, a connection between groups A and B is documented. The entries (A,B) and (B,A) both become one, and the valencies of both A and B are decreased by one. If the set [A,B] does not satisfy one of the constraints, [A,B] is resampled until a set is generated that satisfies all of the constraints. However, if an intermediate molecule is generated that can not have any additional groups added to it, the connection step is restarted and all valencies are reverted to their original values, and all connections are eliminated. When thiophene and aliphatic naphthenic groups are connected, the individual atoms are documented in the connection matrix. The connection algorithm progresses until each and every group is connected to at least one other group.

10. γ-CH_3 carbons are added to any thiophenes, aliphatic associated naphthenic rings, and paraffinic carbons that have valencies greater than zero.
11. The aromatic groups that are connected together by biphenyl bridges are identified. Specific carbons within the aromatic groups that are involved in the biphenyl connections are assigned. The Q_1, Q_2, C_1 and C_2 carbons for each aromatic group are assigned.
12. The aromatic groups, attached to paraffinic groups, thiophenes, and aliphatic naphthenic rings are identified. Two "other aliphatic" carbons are then connected between each aromatic, and its attached group. One of these "other carbons" represents the α methylene carbons that have been discussed in previous steps. The structural and connection matricies are expanded to include these other aliphatic carbons. The specific aromatic atoms involved in these attachments are identified and documented.
13. The aromatic groups with α-CH_3 terminal carbon attachments are identified. The structural and connection matricies are then expanded to include the α-CH_3 carbons. The specific aromatic atoms involved in these attachments are identified and documented.

14. The aromatic groups with β-CH_3 chain attachments are identified. The structural and connection matricies are then expanded to include the beta carbon attachment. The specific aromatic atoms involved in these attachments are identified and documented.
15. Aromatic groups with naphthenic rings are identified. The structural and connection matricies are then expanded to include the four carbons that form each naphthenic ring. For each naphthenic ring, the pair of aromatic atoms that are directly connected to it are identified and documented.
16. Some paraffin chains that are attached to aromatic groups with naphthenic rings, are randomly reattached to the naphthenic rings. The connection matrix is modified to reflect these changes.
17. Both aromatic associated and aliphatic naphthenic rings are identified. Single naphthenic rings are randomly selected and expanded into two ring systems. When a second naphthenic ring is added to another naphthenic ring, the structural and connection matricies are expanded to include the new second ring. The rows in the structural matrix that correspond to carbons of the original naphthenic ring are modified to reflect the changes in the carbon assignments that arise when the second ring is added.
18. α-CH_3 terminal carbons are randomly added to both aromatic associated and aliphatic naphthenic rings. The structural and connection matricies are expanded to include these naphthenic attached α-CH_3 carbons.

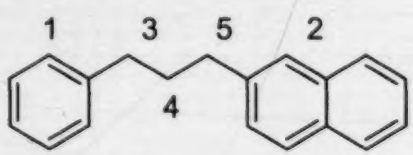
Appendix C. Molecule Storage and Visualization

C.1. Molecule Storage

When it is decided that molecular representations for a petroleum data set are going to be created, a storage framework must be chosen to represent molecules in a computer. First and foremost, the storage framework has to be capable of documenting the types of molecules that are being studied. The chosen storage framework must be capable of documenting and storing: all important connections within a molecule, any topological information that is important, and atomic charges. It must be decided how many molecules the molecular representation will contain and what purpose the molecules will be used for. The chosen framework must be capable of storing the desired number of molecules, given the computing power and memory available. Intuitively, simulations with more molecules will take longer and require more memory than simulations with fewer molecules. Additionally, the chosen storage framework must allow the molecules in the molecular representation to be analyzed and characterized in a desired manner. Furthermore, it must be possible to utilize molecules in the types of simulations that inspired the molecular representations to be created in the first place.

In this study it was necessary to calculate the following properties from the created molecular representations: molecular weights, atomic percentages, and group concentrations. A future objective of this study was to react molecular representations in a thermal cracking simulation. Thus, the chosen storage framework had to be capable of facilitating the development of and being used in thermal cracking simulations.

After careful consideration, a dual matrix storage framework was chosen (Figure C-1). In this framework, a petroleum molecule is composed of a collection of groups. Groups can be defined as individual atoms - such as carbon or sulfur - or as a "primary ensemble" of atoms - such as aromatic and naphthenic rings. The structural matrix (J in Figure C-1) records a description of each group in a molecule. Each row in a structural matrix corresponds to a specific group in a molecule. The columns in a structural matrix are descriptors: each column corresponds to a specific descriptor. Thus, if a molecule contains "n" different groups and if "m" different group descriptors are used to describe a molecule, the structural matrix for the molecule will have n*m dimensions. The individual matrix elements in the structural matrix are 0s and 1s. A row for a specific group is composed entirely of 0s, except for 1s in the columns that contain descriptors of the group.



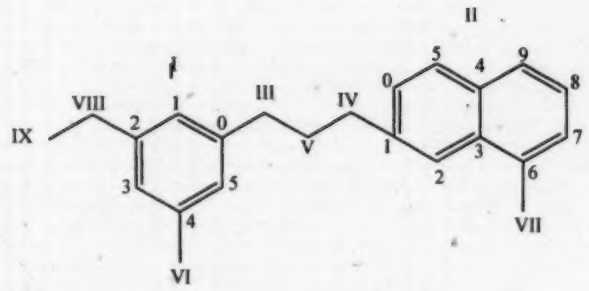
$$J=\begin{array}{cc} \begin{matrix} a & b & c & d \end{matrix} & \\ \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{matrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{array} \qquad C=\begin{array}{cc} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} & \\ \begin{matrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{matrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} \end{array}$$

Figure C-1. The matrix representation of a small hydrocarbon molecule. Matrix J is a structural matrix. Matrix C is a connection matrix. The molecule is represented by five structural groups. Group 1 is a benzene ring while group 2 is a naphthalene ring system. Groups 3 and 5 are α to aromatic aliphatic carbons, while group 4 is a β to aromatic aliphatic carbon. In the connection matrix, column "a" corresponds to benzene rings, "b" corresponds to naphthalene ring systems, column "c" corresponds to α to aromatic aliphatic carbon, while column "d" corresponds to β to aromatic aliphatic carbon.

The connection matrix (C in Figure C-1) documents all of the connections within a molecule. When, a group is denoted as group number "i", it will be described in the i^{th} row of the structural matrix. The connections associated with group number "i" are documented in the i^{th} row and i^{th} column. Thus, the connection matrix is explicitly linked to the structural matrix; the i^{th} row in a connection matrix corresponds to the i^{th} row in the paired structural matrix. Furthermore, the i^{th} column corresponds to the i^{th} row in the connection matrix. A bond connection between group "i" and group "j" is represented by a nonzero entry appearing in both matrix elements (i,j) and (j,i). A zero in a connection matrix for element (i,j) indicates that no connection exists between group "i" and group "j". Thus, the connection matrix is symmetrical about its diagonal.

A problem is created by defining all of the groups as individual carbons except for the aromatic rings - that are documented as primary ensembles of atoms. The problem is that the specific aromatic carbons that are connected to attached groups must all be explicitly specified. If they are not specified, it will not be possible to react molecules in a thermal cracking simulation since the overall structural organization of large petroleum molecules significantly influences how they break apart after a cracking reaction. Also, unless these specific carbons are specified, it would not be possible to develop an algorithm that converts the information stored in the connection and structural matrices into a form that can be visualized.

The solution to this problem is straightforward: the structural matrix is expanded to include an aromatic carbon descriptor column (Figure C-2). Every atom in the aromatic groups is given a number. When a group (i) is connected to an aromatic group (a), the specific aromatic carbon (y) number involved in the connection is documented in the last column of row "i" in the structural matrix.



J=

a	b	c	d	e	f	g	h	
1	0	0	0	0	0	0	0	I
0	1	0	0	0	0	0	0	II
0	0	1	0	0	0	0	0	III
0	0	1	0	0	0	0	0	IV
0	0	0	1	0	0	0	1	V
0	0	0	0	1	0	0	4	VI
0	0	0	0	1	0	0	6	VII
0	0	0	0	0	1	0	2	VIII
0	0	0	0	0	0	1	0	IX

C=

I	II	III	IV	V	VI	VII	VIII	IX	
0	0	1	0	0	1	0	1	0	I
0	0	0	1	0	0	1	0	0	II
1	0	0	0	1	0	0	0	0	III
0	1	0	0	1	0	0	0	0	IV
0	0	1	1	0	0	0	0	0	V
1	0	0	0	0	0	0	0	0	VI
0	1	0	0	0	0	0	0	0	VII
1	0	0	0	0	0	0	0	1	VIII
0	0	0	0	0	0	0	1	0	IX

Figure C-2. A more advanced matrix representation of a small hydrocarbon molecule. Matrix J is a structural matrix. Matrix C is a connection matrix. The molecule is represented by seven structural groups. Group I is a benzene ring while group II is a naphthalene ring system. Groups III and IV are α to aromatic-terminal CH_3 carbons, while group V is a β to aromatic-aliphatic carbon. Groups VI and VII are α to aromatic-terminal CH_3 carbons. Group VIII is an α to aromatic CH_2 carbon that is attached to a β to aromatic carbon- terminal CH_3 carbon (group IX). In the structural matrix, column "a" corresponds to benzene rings, "b" corresponds to naphthalene ring systems, column "c" corresponds to α to aromatic aliphatic carbon, column "d" corresponds to β to aromatic aliphatic carbon, "e" corresponds to α to aromatic terminal CH_3 carbon, while "f" corresponds to α to aromatic CH_2 carbon that is attached to a terminal β to aromatic carbon, "g" corresponds to β to aromatic terminal CH_3 carbon, "h" is the aromatic position holder. The numbers around the aromatic rings are reference positions for aromatic attachments.

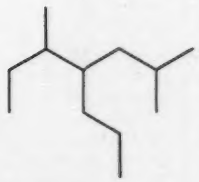
C.2. Molecule Visualization

To represent asphaltene and high molecular weight residue molecules, it is usually necessary to use structural and connection matrices that contain more than 100 rows. Thus, a single heavy oil connection matrix will require at least 100*100=10000 matrix elements. If one has the time and the required motivation and patience, without the aid of a visualization algorithm, it takes a few hours to draw an asphaltene or residue molecule that is represented using the structural and connection matrix methodology. During the course of a thesis that involves creating molecular representations, it is logical that several hundred, if not several thousand molecules will be created. In this study, it was quickly realized that a visualization algorithm must be created.

The goal of the visualization algorithm was to convert information embedded in the connection and structural matrices into a visualization file format that could be easily visualized. Chemical drawing packages currently utilize many different visualization file formats including: MOL, SMILES, CHARMM, and XML. These visualization file formats differ in both technical complexity and in the level of the level of detail that they are capable of describing molecules. Also, these visualization files use different frameworks to represent molecules. For example, MOL files use two matrices to represent a molecule: one matrix to document the (x,y) coordinates of atoms within a molecule and a second matrix to record all of the connections involved within a molecule. Conversely, SMILES files only document the connections involved within a molecule and do not document the coordinates of the atoms of a molecule. Also, some visualization file formats such as SMILES are not easily capable of representing either large or highly branched molecules. In this study, CHEMDRAW was the software package that was to be used to draw and visualize molecules. Thus, the chosen file format had to be compatible with CHEMDRAW. A minimal number of transformation steps were required to

convert the information found in the structural and connection matrices into a visualization file. Thus, a visualization file format that used a connection framework was wanted.

Many visualization file options were considered. The MOL file framework was chosen as the type of file to be used to store the molecules in this study. A MOL file has three parts: a header block, an atom block and a connection block. The header block contains the file name, the total number of atoms in the molecule and the total number of unique connections in the molecule. The atom block specifies the (x,y) coordinates, symbols, mass differences, charges, stereochemistry, and associated hydrogens of each atom in a molecule. The bond block documents the atoms involved, bond type, stereochemistry and topology of each bond in a molecule. A simple molecule and the associated MOL file are presented in Figure C-3.



```
samplemol.mol
 ChemDraw08310622492D

 12 11  0  0  0  0  0  0  0  0999 V2000
-3.2475   -0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
-3.2475    0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
-1.9475    1.5000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
-0.6500    0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 0.6500    1.5000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
-1.9475    3.0000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 1.9475    0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 3.2475    1.5000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 1.9475   -0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
-0.6500   -0.7500    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 0.6500   -1.5000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
 0.6500   -3.0000    0.0000 C  0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  1  0  0  0  0
  3  4  1  0  0  0  0
  4  5  1  0  0  0  0
  3  6  1  0  0  0  0
  5  7  1  0  0  0  0
  7  8  1  0  0  0  0
  7  9  1  0  0  0  0
  4 10  1  0  0  0  0
 10 11  1  0  0  0  0
 11 12  1  0  0  0  0
M  END
```

A

B

C

Figure C-3. A MOL file for a simple hydrocarbon molecule. A. Header Block B. Atom Block: The . C. Bond Block.

A. The file name is specified in the first line. The second line is a descriptor. The first element in the third line specifies the number of atoms in the molecule. The second element specifies the number of bonds in the molecule.
B. The first three columns in the Atom block specify the (x,y,z) coordinates of each atom. The fourth column documents the symbol of each atom. In ChemDraw, hydrogen is implicit. Therefore MOL files that are used by

ChemDraw, do not require that hydrogen be explicitly defined. The other columns define the mass differences, charges and stereochemistry associated with each atom.

C. In a specific row, the atoms involved in a bond are documented in the first two rows. The bond type is documented in the third column: 1 represents a single bond. 2. represents a double bond. The other columns are used to define the stereochemistry and topology of the each bond.

The connection matrix documents all of the carbon-carbon and carbon-sulphur bonds within a molecule, except for those within aromatic groups. However, the structural matrix documents all of the groups that compose a molecule. Thus, both types of matricies must be used when creating a MOL file using the visualization algorithm. The sequential steps that the created visualization algorithm utilizes are:

1. The connection matrix is used to generate a nearly complete bond block that contains aromatic groups. During this step, the structural matrix is used to identify aromatic groups. The visualization algorithm inserts the aromatic groups into the bond block.
2. The connection matrix is used to generate a complete atom block- using (x,y) coordinates- that includes information for the atoms found within the aromatic groups. During this step, the structural matrix is used to identify atomic symbols. The visualization algorithm inserts rows to describe the aromatic atoms.
3. The visualization algorithm uses the connection matrix to identify naphthenic rings. In stage one, single bonds within the naphthenic rings are sometimes missed. However, all of the naphthenic atoms are identified in stage two. In Stage 3. the visualization algorithm expands the bond block to include all missing naphthenic bonds.
4. The visualization algorithm reassigns the atoms involved in aliphatic-aromatic bonds to be consistent with the ordered assignments that are recorded in the structural matrix.

The linear algebra that was used within the visualization algorithm utilizes the vector properties of the connection matrix. The connection matrix documents the connections of a specific group to each and every other group within a molecule. Thus, each row in a connection matrix is a vector that documents the spatial connections to each and every other vector. If the connection matrix has dimensions c*c, a vector corresponding to an individual group will have dimensions 1*c-with a total of c elements. If position "r" in vector c is a nonzero element, groups "c" and "r" are connected.

Vector addition is the linear algebra method behind the visualization algorithm used in this study. Two vectors are necessary to drive the algorithm: an addition vector (**a**) that is sequentially added to the rows/vectors found in the connection matrix, and a holder vector (**b**) that identifies new groups at each step. At each step, (**b**) is used to identify all of the groups that have not been analyzed. The previous value of (**a**) is compared to the current value of (**a**). When a group "q" is present in the current value of (**a**) but not present in the previous value of (**a**), the holder vector (**b**) is assigned a value of 1 at position "q". When a group "m"-with corresponding vector (**m**) is identified as not having been previously analyzed, a transformed vector (**mt**) is created that has all of the elements of (**m**) but has the numerical values of the elements of (**m**) replaced with "m". (**mt**) is then added to (**a**). Thus, a numerical value "nw" stored in (**a**) at position "w", corresponds to an adjacent group that is directly connected to group "w". At each step, (x,y) coordinates are assigned to each group identified using vector (**b**). The (x,y) coordinates of group (**q**) are calculated by adding a small incremental amount to the (x,y) coordinates assigned to the adjacent group-"nw"-that is present in **a**(q). The (x,y) coordinates are then documented in the bond block of the molecule. The sequential addition progresses until all of the atoms in the molecule are taken into account.

Appendix D. Modeling of Thermal Cracking Reactions

The dual structural-connection matrix methodology discussed in Appendix B.1. explicitly defines a petroleum molecule; the individual atoms and groups that compose a molecule are defined in the structural matrix while the connection matrix documents how these groups and connected together. These two matrices contain sufficient information to extract the structural information necessary to react a molecule in a thermal reaction simulation. Thus, an algorithm can be developed that analyzes the connection and structural matrices of a molecule and calculates all of the necessary structural information necessary for thermal cracking simulations based on model compound data. The nonlinear nature of the molecular structural arrangement of residue molecules renders extracting the necessary information to be challenging.

D.1. Incorporation of Model Compound Data

The model compound cracking studies discussed in Appendix A allow the cracking probabilities of C-C and C-S bonds to be determined by their proximity to key structural groups. The key structural groups are: aromatic rings, pyrene rings, naphthenic rings, thioethers, and chain methynes. The functional groups described in the model compound studies are often located within a single petroleum molecule. Thus, there was a requirement to express the data for each model compound relative to a common standard – a reference bond (or reference bonds). A reference bond is defined as a bond that is at least eight carbons away from a key structural group. Since a reference bond is far away from a functional group, the group's influence on the thermal cracking of the reference bond is very minimal. For each model compound study, the reaction products were first defined by the distance between the key structural group and terminal C-C bond in the product. For example, in the case of phenyldodecane

thermal cracking; phenyldodecane is classified as having eleven carbons between the phenyl ring system and the terminal side chain carbon (Figure D-1).

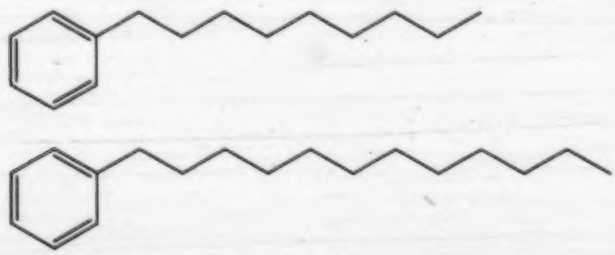


Figure D-1. The structure of a. phenynonane b. phenyldodecane.

For the cracking of phenyldodecane, the concentrations of all the thermal products were normalized relative to the average concentrations of reaction products with side chains with lengths of eight (phenyloctane and phenyloctene), nine (phenylnonane and phenylnonene) , and ten carbons (phenyldecane and phenyldecene). Likewise, for each of the other model compounds, the reaction products were normalized relative to the average concentrations of the reaction products that have chain lengths of eight, nine, and ten relative to the key structural group. After the normalization, olefin products were lumped together with the other alkane products. The resulting normalized and lumped concentrations are referred to as relative cracking propensities.

D.1.2. Relative Cracking Propensities

A relative cracking propensity represents the likelihood of cracking a specific bond compared to the likelihood of cracking a reference bond. Since the reference bonds (those bonds that are eight, nine and ten carbons away from a key structural group) are the same for all of the model compounds, this approach allows all of the model compound data to be used for a single molecular representation molecule. For example, the relative cracking propensities of the

aromatic model compounds, yield relative cracking probabilities at different C-C positions relative to a core aromatic group.[1,2] Likewise, the relative cracking propensities of tridecylcyclohexane yield the cracking probabilities at different C-C positions relative to a central naphthenic ring[3]. In this study, it will be assumed that the same cracking chemistry occurs with hydroaromatics, and thus the relative cracking propensities calculated from the tridecylcyclohexane study can be applied to hydroaromatics. However, the relative cracking propensities can not be as easily applied to the thioether model compound, didoceyl sulfide. In this study, only C_1-C_{12} olefin products were observed. Thus all of the observed products resulted by the cracking of the C-S bond. The relative cracking propensity associated with the C_{12} product will be indicative of the relative reactivity of the C-S bond. However, the relative cracking propensities associated with the other products, do not necessarily yield the cracking probabilities of the C-C bonds relative to the C-S bond since these products might have formed after the cracking of a C-S bond and possibly after the liberation of the S from the C_{12} fragment. However, in this study, the relative cracking propensities calculated from didoceyl sulfide will be used to calculate the C-C bond cracking probabilities relative to thioethers.[4] The probabilities of scission reported by Kissin using isoalkane model compounds, were used to calculate the relative cracking propensities of CH-CH_2 bonds and the adjacent CH_2-CH_2 bonds.[5] The relative cracking propensities used in this study are presented in Figure D-2.

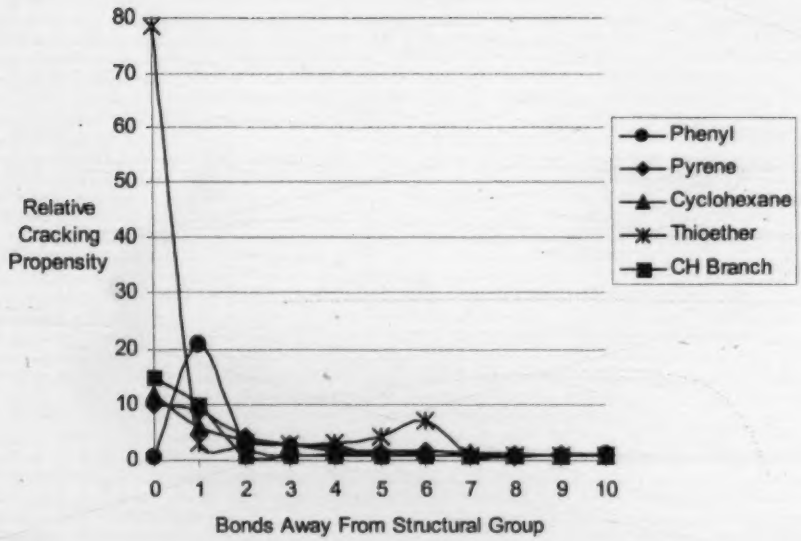


Figure D-2. Relative cracking propensities of various C-C and C-S bonds in a variety of model compounds. The following respective model compounds and literature references were used to construct this plot: •Phenyl Phenyldodecane[1], ♦Pyrene-1,20-di(1-pyrenyl)Eicosane[2], ▲Cyclohexane-Tridecyl Cyclohexane[3], * Thioether- Didodecyl Sulfide[4], ■CH Branch- Pristane, Phytane, Squalane.[5]

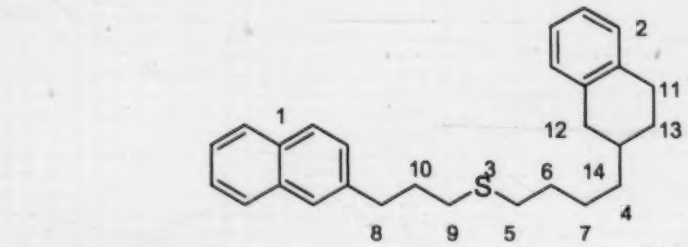
When all of the thermal cracking products for all of the model compounds are normalized in such a manner, the result is a detailed series of relative propensities that give measurements of the likelihood of various thermal products being produced. Thus, the first step in determining the relative cracking propensities for a specific bond in a molecule, is to determine the key groups that are nearest the bond. The second step involves determining the number of C-C bonds between the specific bond in question and the key groups. Once these distances are determined, the relative cracking propensities (based on the nearest key groups) of a specific bond can be estimated from the normalized model compound cracking data. When a C-C bond is directly adjacent to a key group, the normalized model compound data for that key group is used to determine the relative propensity of cracking for that bond. However, many C-C

bonds are near two key groups. Thus, the relative cracking propensity of a C-C bond will have different values, depending on which key group is used for the relative propensity assignment. In the cases when two relative cracking propensities are obtained, the larger value of the two was used.

D.2. Thermal Cracking Modeling

D.2.1. The REACT Matrix

Each reactive C-C and C-S bond is documented in a matrix called the REACT matrix (Figure D-3). The REACT matrix documents the nearest key groups to a specific bond, the type of key groups, and the distances of the C-C or C-S bond from the key groups. The REACT matrix does not document the C-C bonds that are nonreactive, such as aromatic C-C bonds. An explicit description of the REACT matrix framework is presented in Figure D-3. The REACT matrix essentially documents all reactive connections in a molecule and the distance from important reactive groups. The REACT matrix is the essential framework of the thermal cracking algorithm. The REACT matrix is used to assign cracking propensities to each reactive bond.



Connections

Type of Key Groups
Nearest Bond

Type of Key Groups
Nearest Bond

Distances From
Key Groups

1	8	1	2	1 4	1 3
8	10	1	2	2 3	1 3
10	9	1	2	3 2	1 3
9	3	1	2	4 1	1 3
14	4	3	2	1 5	14 3
4	7	3	2	2 4	14 3
7	6	3	2	3 3	14 3
6	5	3	2	4 2	14 3
5	3	3	2	5 1	14 3
2	11	0	0	0 0	0 0
11	13	0	0	0 0	0 0
13	14	0	0	0 0	0 0
14	12	0	0	0 0	0 0
12	2	0	0	0 0	0 0

Figure D-3. The REACT matrix. Columns one and two document the specific connections within a molecule. For illustrative purposes, column one will document group "A", while column two will document group "B". Column 3 documents the type of key group (key group A) that is nearest to group A (the key groups are defined in Table 3-10). Column 4 documents the key group (key group B) that is nearest to group B. Column 5 documents the distance (number of C-C or C-S bonds) between bond A-B and key group A. Column 6 documents the distance between bond A-B and key group B. Column 7 specifically defines the group that represents key group A Column 8 specifically defines the group that represents key group B.

D.2.2. Thermal Cracking Simulations

The reactions modeled in this study are presented in Figure D-4. Reaction 1 represents the breakage of a C-C bond while reaction 2 represents the breakage of a C-S bond. The relative cracking propensities for these reactions are presented in Figure D-2. Reaction 3 represents the dehydrogenation of naphthenic rings to aromatic rings. When a molecule was stochastically chosen to crack, its napthenic rings were stochastically sampled to change to aromatic rings. Each naphthenic ring was given a 50% chance of conversion. Only one naphthenic ring in two ring systems was allowed to convert to an aromatic ring.

When a single bond is broken (excluding bonds within naphthenic or primary order cyclic organizations) in a molecule, two product molecules are formed. Each product molecule has unique connection and structural matrices. All of the connection information in the product matrices originated from a reactant connection matrix. The same vector addition principles that were used to create the visualization algorithm were used to construct an algorithm that constructs connection matrices for reaction products. A flow diagram of the thermal cracking algorithm is presented in Figure D-5. A detailed description of the thermal cracking algorithm is presented in Appendix D.3. Essentially, at each reaction step, the relative cracking propensities were used to randomly sample a series of molecules. The thermal cracking simulations were divided into reaction steps. At each reaction step, ten molecules were stochastically chosen to crack. The higher a relative cracking propensity, the higher the likelihood of cracking a bond. The relative cracking propensities were then used to stochastically sample single bonds within each sampled molecule to crack. Each feed molecule could react multiple times and generate multiple products. Likewise, the product molecules that were created at various reaction steps could react further, as long as they were sampled during a reaction step.

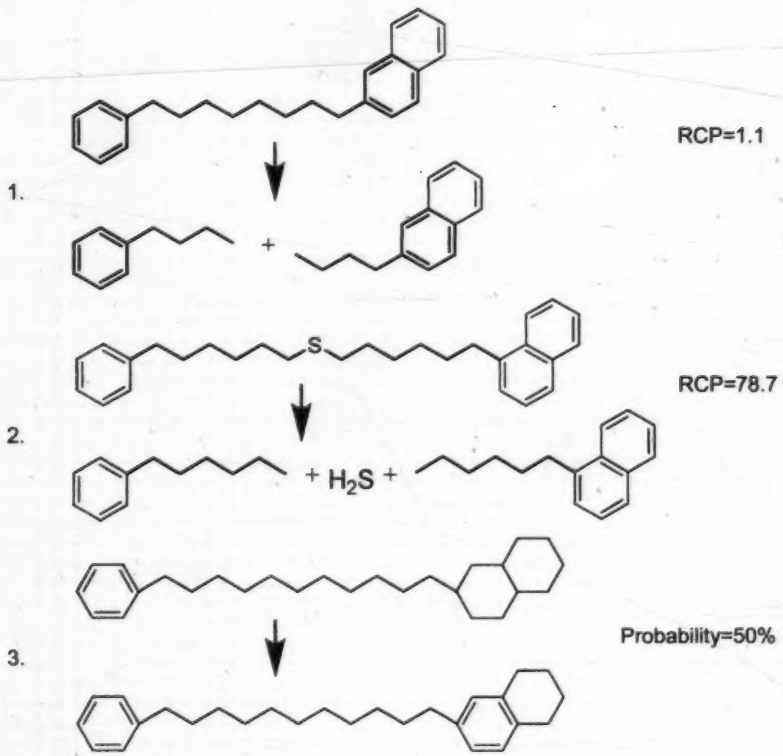


Figure D-4. Transformations and reactions modeled in the thermal cracking simulations. 1. C-C bond cracking. 2. C-S bond cracking. 3. Dehyrdogenation of naphthenic rings to aromatic rings. RCP represents relative cracking propensity.

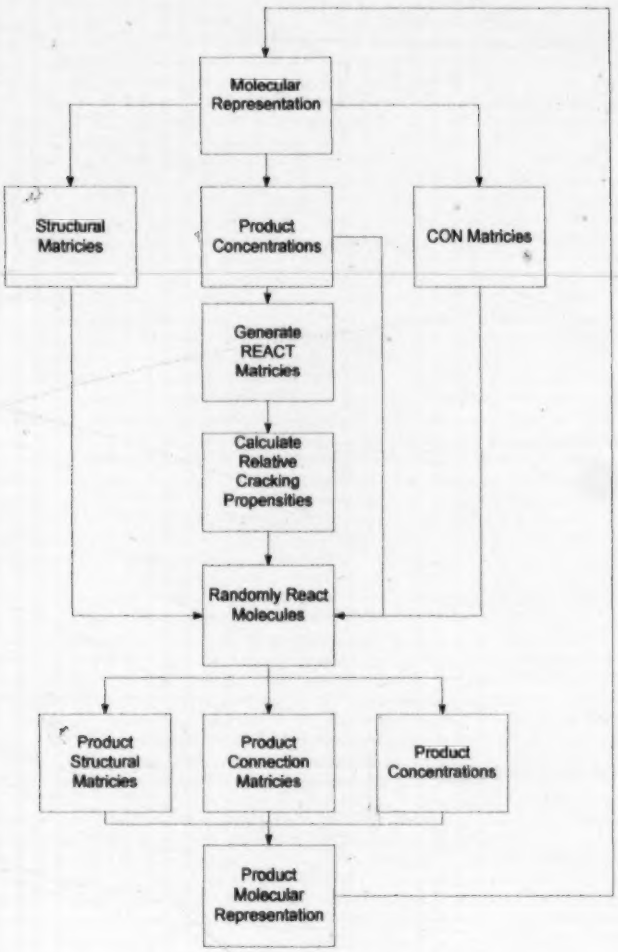


Figure D-5. Overview of the thermal cracking algorithm.

D.3. The Thermal Cracking Algorithm

When a single bond is broken (excluding bonds within naphthenic or primary order cyclic organizations) in a molecule, two product molecules are formed. Each product molecule has unique connection and structural matrices. All of the connection information in the product matrices originates from the reactant matrix. The same vector addition principles that were used to create the visualization algorithm were used to construct an algorithm that constructs connection matrices for reaction products. After a specific bond is broken, a vector (**a**) is defined as being the first row in the connection matrix. At each step, all nonzero elements in (**a**) are identified, and the corresponding rows in the connection matrix are added to (**a**). When no more unique elements can be added to (**a**) the sequential vector addition is stopped. At this point (**a**) contains a mapping of all the groups contained within one of the two product molecules. (**ONE**) will be defined as a vector that is the same size as (**a**) and contains 1 in every vector element. When (**a**) is subtracted from (**ONE**) the resulting vector (**b**) contains a mapping of the second molecule. (**a**) is first used to create structural and connection matrices for the first product molecule. When an element in (**a**) is nonzero, this indicates that a group from the original molecule is present in this particular product molecule. Conversely, when an element in (**a**) is zero, this indicates that this group is absent from this particular product molecule. (**b**) is then used to create structural and connection matrices for the second product molecule. The resulting product connection matrices have entire rows and columns filled with zero elements, since the dimensions of the product connection matrices will be initially the same as the original reactant molecule. The connection matrices are then scanned and the rows and columns that only contain 0 elements are deleted.

The REACT matrix is the essential framework of the thermal cracking algorithm. For each molecule in a molecular representation, a REACT matrix

documents the reactive bonds and their proximity to key structural groups that have a key role in the reaction chemistry. The information in the REACT matrix is used to assign "cracking propensities" to each bond in a molecule. "Cracking propensities" are relative likelihoods of a bond breaking during a cracking step. The following is a brief description of the important steps in the cracking methodology used developed in this thesis:

1. A REACT matrix is created for each molecule either during the construction process or after – using the structural and connection matrices. Every molecule is represented by a connection and structural matrix. Each REACT matrix is sorted so that:
    a. All of the rows documenting connections between two key groups (A and B) are grouped together.
    b. Each ordered grouping of connections (from 1a.) is sequentially arranged according to the relative distances from key groups A and B. For example, a given row "f" in the connection matrix will document a connection that is "q" bonds away from group A and "w" bonds away from group "a". Rows "e" and "g" will represent rows that respectively are above and below row "g" in the REACT matrix. The bond that is documented in row "e" will be one bond closer to group A, but one bond further away to group B, compared to the bond documented in row "f". Likewise, the bond that is documented in row "g", will be one bond further away to group A, but one bond closer to group B, relative to the bond documented in row "f". In turn, the bond that is documented in row "g", will be two bonds further away to group A, but two bonds closer to group B, relative to the bond documented in row "e".
    c. For all adjacent rows in REACT, "c" and "d" (with "d" appearing in a row directly below row "c") that document a common group "u" that is found either in column 1 or 2, the connections are

arranged so that group "u" appears in column two of row "c" and column 1 of row "d".

2. Using the REACT matrix, the relative cracking propensity (RCP) of each bond in a molecule are calculated using the model compound data from Figure D-2. Columns 5 and 6 are used to determine the distance from the key groups defined in columns 3 and 4. The cracking propensity of a bond relative to each key group is then determined relative to each key group using the model compound data in Figure D-2. For a given bond, the greater of the two values is assigned as the relative cracking propensity.

   At each reaction step, the REACT matricies for each molecule are analyzed and the summation of the RCPs (relative cracking propensities) for each molecule are calculated. For each molecule, the RCP summation is multiplied by the integer molecular composition of the molecule. These resulting products are referred to as mole weighted RCPS (MWRCPs).

3. At each reaction step, ten molecules are stochastically sampled using the MWRCPs. At each reaction step, the REACT matricies are reformulated. Thus as a feed molecule reacts, the RCP for a specific bond will change as its surrounding groups react and chemically change.
4. Absolute cracking probabilities for each reactive bond in a reacting molecule, are calculated by dividing each relative cracking propensity by the summation of all relative cracking propensities of the molecule.
5. The bonds in the REACT matrix for each reacting molecule chosen in step 3. are stochastically sampled and a single bond is chosen to crack during each cracking reaction. The concentration of a reacting feed molecule is decreased by 1 molecule, while each product has a resulting composition of 1 molecule.

6. The naphthenic ring systems of all molecules are stochastically sampled. 50% of the sampled naphthenic ring systems are converted to aromatic rings. In two ring naphthenic ring systems, only one of the two rings is allowed to transform into an aromatic ring.
7. For each bond that cracks, two molecules are generated. The connection and structural matrices of each product are generated and stored in arrays. Before the product matrices are created, the groups near the bond (A,B) that is being cracked are analyzed and if necessary, their documentation in the structural matrix is modified. Columns 1,2, 5 and 6 of the REACT matrix are used to identify all of the groups between group A and the nearest key group to A, and between group B and the nearest key group to B. The groups that are identified are then reassigned in the structural matrix. This assures that all groups in the product molecules are consistent with their original NMR spectroscopy definitions.
8. The naphthenic and aromatic contents of the cracked molecules are analyzed. Those molecules that do not contain any aromatic or naphthenic carbons, and that contain fewer than ten carbons in total, and do not contain N or S, are classified as gas molecules and removed from the liquid fraction.
9. The matricies of the liquid product molecules are combined with the matricies of the reacted feed molecules. REACT matricies are then created for the product molecules.
10. The molecules continuously react, the product molecules are lumped in with the input feed molecules, and REACT matricies are produced for new product molecules, until either:
    a. A given number of reaction steps is reached.
    b. A given conversion or product composition is obtained.
    c. The computer memory becomes full.

D.5. References

1. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1987**, *26(2)*, 374-376.
2. Freund, H.; Matturro, M.G.; Olmstead, W.N.; Reynolds, R.P.; Upton, T. H. *Energy & Fuels.* **1991**, *5(6)*, 840-846.
3. Savage, P.E.; Klein, M.T. *Ind. Eng. Chem. Res.* **1988**, *27*, 1348-1356.
4. Abikhers, V.; Fixari, B.; Leperchex, P. *Fuel.* **1986**, *65 (3)*, 442-446.
5. Kissin, Y.V. *Ind. Eng. Chem. Res.* **1987**, *26 (8)*, 1635-1638.

Appendix E. The Second Molecule Construction Algorithm

This Appendix provides a detailed description of the second molecular representation creation algorithm presented in Chapter 5.

The NMR spectroscopy data used in Chapter 5 required a new method of molecular construction to represent all of the groups that it resolved. The methods presented in this Appendix used the same sampling and construction methods as the methods outlined in Appendix C. This construction method was also created to facilitate and minimize the required effort of developing the thermal cracking and REACT creation algorithms. This new algorithm was designed to be more computationally efficient. Specifically, the number of random connection steps was significantly decreased. The algorithm described in Appendix C randomly connects aromatic groups to various aliphatic species, including various CH_2 groups. Conversely, the algorithm in this section randomly connects aromatic groups, thioethers, and aliphatic naphthenic groups together, and then connects CH_2 groups in between the resulting pairs of groups. The following is a description of the steps involved in this algorithm.

Phenylnaphthalene and biphenylnaphthalene were not included in this algorithm, since it was capable of created aromatic clusters with multiple aromatic groups connected together via biphenyl bridges. Porphyrins and dibenzofurans were not included since oxygen data was not available. Preliminary simulation indicated that the molecular represnetaitons had closer fit to the analytical data when both benzothiophene and dibenzothiophene were included.

1. The concentrations of benzene, naphthalene, phenanthrene, pyrene, benzothiophene, dibenzothiophene, and indole are first optimized with nonlinear optimization using the NMR spectroscopy derived

concentrations for "Aromatic NS (C+CH)", "Aromatic Bridge C", "Aromatic Alkyl Subsituted C", and "Aromatic CH".

2. Aromatic groups are randomly sampled and documented in the structural matrix. The maximum number of attachments (MNA) allowed for each aromatic group type are presented in Table E1.

Table E1. Maximum number of attachments allowed to connect to the aromatic groups.

Aromatic Group	Maximum Number of Attachments (MNA) Allowed
Benzene	3
Napthalene	4
Phenanthrene	5
Pyrene	7
Benzothiophene	3
Dibenzothiophene	4
Indole	3

3. Aliphatic naphthenics, chain methyne, and thioether groups are then randomly sampled and documented in the structural matrix. Each of these groups are allowed to connect to a maximum of two other groups. At this point each of these groups is represented as a single entry in the structural matrix.
4. Aromatic naphthenic groups are randomly sampled and assigned to specific aromatic groups. At this point the naphthenic groups that are attached to aromatics are represented as a single entry in the structural matrix.
5. A vector called HO is created. HO documents the specific groups that have been connected during the construction process. A "1" element in HO means that a group has not yet been connected. A "0" element in HO

means that a group has been connected. Thus, when a group is connected to another group during the construction process, the HO element associated with that group changes from 1 to 0. Thus, when all of the groups for a molecule are connected, the sum of the HO vector for the entire molecule (all of the groups) will be zero.

6. A connection matrix is created that has rows corresponding to: aromatic, chain methyne, and aliphatic naphthenic groups. A connection algorithm randomly connects all of the groups together subject to a few constraints:

   -Groups must only connect to other groups. They are not allowed to connect to themselves.

   -Connections are only allowed between a group that has already been connected to the developing molecule, and a group that does not have any connections.

   -Group valencies are not allowed to proceed below zero. Each group has a predetermined number of attachments that it is allowed to connect to.

The valencies are defined as:

$$v_{aro_i} = MNA_i - \sum attached\ aromatic\ groups - \sum attached\ aliphatic\ naphthenics - \sum attached\ thioethers - \sum attached\ chainmethynes$$

$$v_{alinap} = 2 - \sum attached\ aromatic\ groups - \sum attached\ aliphatic\ naphthenics - \sum attached\ thioethers - \sum attached\ chain\ methynes$$

$$v_{thio} = 2 - \sum attached\ aromatic\ groups - \sum attached\ aliphatic\ naphthenics - \sum attached\ chain\ methynes$$

$$v_{CM} = 2 - \sum attached\ aromatic\ groups - \sum attached\ aliphatic\ naphthenics - \sum attached\ chain\ methynes - \sum attached\ thioethers$$

For v_i:

MNA_i represents the maximum number of attachments that aromatic group *i* is allowed to have.

attached aromatic groups represent aromatic groups that are attached to group "i".

attached aliphatic naphthenics represent aliphatic naphthenic groups that are attached to group "i".

attached thioethers represent thioether groups that are attached to group "i".

attached chainmethynes represent chain methynes groups that are attached to group "i".

v_{aro_i} represents the valency of a aromatic group "i".

v_{alinap} represents the valency of an aliphatic naphthenic .

v_{thio} represents the valency of a thioether.

v_{CM} represents the valency of a chain methyne.

At each step, the connection algorithm randomly samples two groups, A and B. When A and B do not conflict with any of the constraints, a connection between groups A and B is documented. The connection matrix entries (A,B) and (B,A) both become 1, and the valencies of both A and B are decreased by 1. At each stage when an (A,B) connection is documented in the connection matrix, a row (c) is created in the REACT matrix. Group A is documented in the first column of C while group A is

documented in the second column. If the set [A,B] does not satisfy all of the constraints, [A,B] is resampled until a set is generated that satisfies all of the constraints. However, if an intermediate molecule is generated that can not have any additional groups added to it, the connection step is restarted and all valencies are reverted to their original values, and all connections are eliminated. When an aliphatic naphthenic, thiophene, or chain methyne group is connected, the individual atoms are documented in the connection matrix. The connection algorithm progresses until the sum of HO is 0 (each and every group is connected to at least one other group).

7. Terminal γ-CH_3 carbons are then randomly sampled and assigned and connected to available groups. The structural and connection matricies are then expanded and documented accordingly. Rows in the REACT matrix are created to document all of these new connections.
8. The REACT matrix is then scanned to identify the specific connections (anchor groups) documented within it. Connections are classified as:
    a. Aromatic group connected to another aromatic group
    b. Gamma terminal carbon connected to a thioether group
    c. Chain methyne connected to a gamma terminal groups
    d. Alipahtic nahthenic group connected to a gamma terminal carbon
    e. All other connections not defined in a-d. (i.e. aromatic-thiophene, aromatic-aliphatic naphthenic, thiphene-aliphatic naphthenic, aromatic-chain methyne)
9. For all of the connections identified in step 7 linking carbon attachments are randomly sampled and documented in the REACT matrix.
10. A secondary connection algorithm is then used to connect all of the linking carbon attachments that were sampled in step 8 to the anchor groups (identified in step 7) that they are connected between. The steps involved in this algorithm include:
    a. Linking carbons are first directly attached to the anchor groups identified in step 7. These linking carbons are specifically defined

and documented in the structural matrix. The connections associated with these carbons are documented in the connection and REACT matrices.

b. The linking chains are then sequentially expanded until the chain lengths between each pair of anchor groups identified in step 7 are equal to the number of carbons sampled in step 8. Each linking carbon is specifically defined and documented in the structural matrix. The connections associated with each linking carbon are documented in both the connection and REACT matricies.

c. The number of carbons between the linking carbons and the nearest anchor groups are calculated and documented in the REACT matrix.

11. The structural and connection matricies are expanded to include all of the carbons found in the aliphatic naphthenic groups. The connection matrix is scanned and modified to document the rearranged connections that result when the aliphatic naphthenics are explicitly documented. The REACT matrix is expanded to document all of the internal connections found within the aliphatic naphthenic ring systems.

12. The structural and connection matricies are expanded to include all of the carbons found in the thioether groups. The connection matrix is scanned and modified to document the rearranged connections that result when the thioether groups are explicitly documented. The REACT matrix is expanded to document all of the internal connections found within the thioether groups.

13. Aromatic groups with napthenic rings are identified. The structural and connection matricies are then expanded to include all four carbons that form each naphthenic ring. For each naphthenic ring, the pair of aromatic atoms that are directly connected to it are identified and documented in the connection matrix. The REACT matrix is expanded to document all of the connections found within the aromatic naphthenic rings.

14. Both aromatic associated and aliphatic naphthenic rings are identified. Single naphthenic rings are randomly selected and expanded into two ring systems. When a second naphthenic ring is added to another naphthenic ring, the structural and connection matricies are expanded to include the new second ring. The rows in the structural matrix that correspond to carbons of the original naphthenic ring are modified to reflect the changes in the carbon assignments that arise when the second ring is added. The connections within the newly added naphthenic rings are then documented in the REACT matrix.
15. Naphthenic alpha, ethyl and butyl attachments are randomly sampled and connected to naphthenic rings. The sampled groups are documented in the structural matrix. The connection matrix is expanded to include the new naphthenic attachments.
16. For each aromatic group a vector containing a random permutation of all allowable attachment positions is generated. The position of each group that is attached to an aromatic group is first assigned using the randomly generated position vector for that group.
17. Excluding aromatic naphthenic rings, the position vectors generated in step 15 for each aromatic group are used to assign specific ring attachment positions to all groups that are directly attached to aromatic groups.
18. For each aromatic group, the following groups are randomly sampled:
    a. Methyl CH_3 attachments
    b. Ethyl C_2H_5 chain attachments
    c. Propyl C_3H_7 chain attachments
    d. Butyl C_4H_9 chain attachments
    e. Pentyl C_5H_{11} chain attachments
    f. Hexyl C_5H_{13} chain attachments
    g. Octyl C_6H_{15} chain attachments
    h. Nonyl C_7H_{17} chain attachments
    i. Butyl C_8H_{19} chain attachments

j. Decyl C_9H_{21} chain attachments
k. C_3H_7 methyl branched chain attachments
l. C_4H_9 ethyl branched chain attachments
m. C_6H_{13} butyl branched chain attachments
n. C_4H_9 methyl branched chain attachments
o. C_5H_{11} ethyl branched chain attachments
p. C_7H_{13} butyl branched chain attachments
q. C_5H_{11} methyl branched chain attachments
r. C_8H_{13} ethyl branched chain attachments
s. C_8H_{15} butyl branched chain attachments
t. C_6H_{13} methyl branched chain attachments
u. C_7H_{15} ethyl branched chain attachments
v. C_9H_{15} butyl branched chain attachments
w. CH olefin attachments
x. C_4H_7 olefin attachments
y. C_5H_9 olefin attachments

When a group is sampled it is documented in the structural matrix. The connection of the group to the aromatic group is documented in the connection matrix, as well as all connections within the attached group. The REACT matrix documents all connections within the attached groups, in addition to the distance of each carbon in the attached group to the aromatic group.

19. Some aliphatic attachments that are attached to aromatic groups with naphthenic rings, are randomly reattached to the naphthenic rings. The structural, connection and REACT matrices are modified to reflect these changes.
20. Methyl, ethyl and butyl attachments are randomly attached to chain methyne carbons. The connection and structural matrices are expanded and updated accordingly.